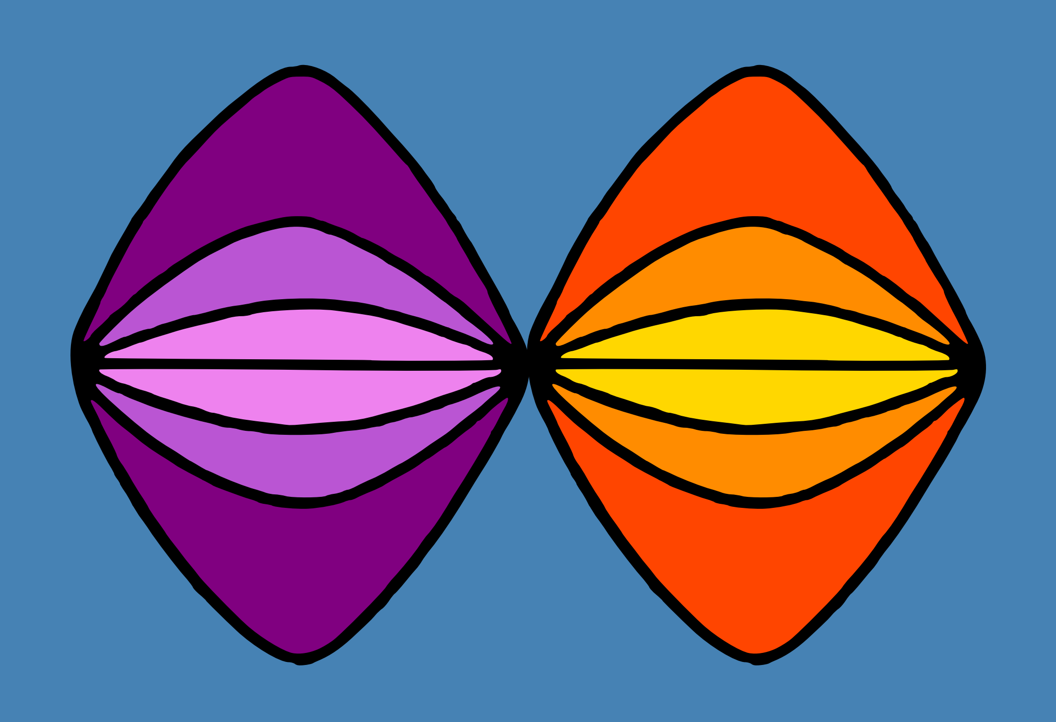
H. Ventsel

THÉORIE DES PROBABILITÉS



Éditions Mir Moscou

Е. ВЕНТЦЕЛЬ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ИЗДАТЕЛЬСТВО «НАУКА» · МОСКВА

H. VENTSEL

THÉORIE DES PROBABILITÉS

ÉDITIONS MIR • MOSCOU

PREMIÈRE ÉDITION 1973
RÉIMPRESSION 1977
2[e] RÉIMPRESSION 1982

TRADUIT DU RUSSE
PAR A. SOKOVA

На французском языке

AVANT-PROPOS

Le présent manuel de la Théorie des probabilités édité en U.R.S.S. en 1962 et réédité, complètement remanié, en 1969, est le contenu d'un cours qui fut enseigné depuis dix dernières années aux élèves des établissements d'enseignement technique supérieur de Moscou.

L'auteur s'est fixé pour but d'exposer le sujet le plus simplement possible et n'a pas jugé utile de se conformer strictement à la trame d'une analyse mathématique rigoureuse.

Certains théorèmes sont présentés sans démonstration : exposé sur les limites de confiance et les probabilités de confiance, théorème de Kolmogorov sur critère de conformité.

L'appareil mathématique utilisé ne réclame pas de connaissances en mathématiques ne figurant pas aux programmes des Ecoles techniques. Là où des notions nouvelles sont introduites (opérateurs linéaires, matrices, formes quadratiques, etc.) on fournit des explications nécessaires.

Pour faciliter l'assimilation des différents chapitres du cours l'auteur a ajouté à l'exposé de nombreux exercices d'application avec les solutions.

L'auteur sera très reconnaissant au lecteur pour toute remarque ou suggestion sur le contenu ou la présentation de cette édition française de l'ouvrage.

Hélène Ventsel
le 21 avril 1973

CHAPITRE PREMIER

INTRODUCTION

1.1. Objet de la théorie des probabilités

La théorie des probabilités est une science mathématique étudiant les lois régissant les phénomènes aléatoires. On dit qu'un phénomène est *aléatoire* si, reproduit maintes fois, il se déroule chaque fois un peu différemment, de sorte que le résultat de l'expérience change d'une fois à l'autre d'une manière aléatoire, imprévisible.

Voici quelques exemples de phénomènes aléatoires.

1. Une expérience consiste à jeter un dé (un cube à six faces portant un, deux, trois, quatre, cinq ou six points) (fig. 1.1.1).

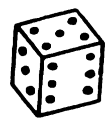


Fig. 1.1.1

Le résultat de l'expérience est le nombre de points amenés par le dé. En répétant l'expérience plusieurs fois on verra son résultat changer d'une manière aléatoire imprévisible. Ces différences sont liées à l'action de divers effets, parfois peu importants et même imperceptibles, intervenant chaque fois que l'on jette le dé et que celui-ci tombe sur la table, comme, par exemple, l'impulsion initiale communiquée au dé, les rugosités de la surface de la table, les vibrations du plancher sur lequel se trouve la table, etc.; même lorsque toutes les mesures nécessaires à la conservation des conditions de l'expérience sont prises, on ne peut assurer la stabilité des résultats.

2. Une expérience consiste en ce que la caisse automatique d'un autobus est remplie de pièces de monnaie durant toute la journée. On vide cette caisse en fin de journée et l'on compte la recette (résultat de l'expérience). Il est clair que ce résultat varie d'un jour à l'autre et oscille autour d'une valeur moyenne dans certaines limites. Ces variations sont dues à des facteurs aléatoires tels que le nombre de voyageurs montés dans l'autobus, le nombre de voyageurs ayant des billets hebdomadaires ou n'en ayant pas, etc.

3. Une expérience consiste à peser un certain nombre de fois un corps sur une balance automatique. Les résultats des expériences répétées ne sont pas rigoureusement constants. Les différences sont liées à l'action de nombreux facteurs, comme, par exemple, la position du corps sur le plateau de la balance, les vibrations aléatoires de la table, les erreurs d'estimation des indications de l'appareil, etc.

4. Une expérience consiste en ce qu'un avion effectue un vol régulier entre deux points. Le temps de vol n'est pas constant d'une expérience à l'autre mais varie quelque peu. Ces variations (le plus souvent ce sont des retards) sont liées à l'influence des conditions météorologiques sur la ligne aérienne, aux défauts d'appareillage, etc.

5. On tire d'un engin disposé sous un angle θ par rapport à l'horizontale des obus à une vitesse initiale v_0, les caractéristiques balistiques des obus (poids, forme) étant les mêmes. La trajectoire théorique est entièrement déterminée par ces conditions, la trajectoire expérimentale s'en écartera obligatoirement à cause des effets secondaires, comme les erreurs de fabrication de l'obus, l'écart du poids de sa valeur nominale, l'erreur de mise en place de l'engin, les conditions météorologiques, etc. Si plusieurs coups sont tirés dans des conditions à première vue similaires, on obtient non pas une seule courbe théorique, mais tout un faisceau de trajectoires, donnant la dispersion des obus.

Tous les exemples mentionnes ci-dessus sont envisagés d'un même point de vue faisant ressortir les variations aléatoires, les résultats différents d'une série d'expériences dont les conditions essentielles sont maintenues constantes. Toutes ces variations sont toujours liées à la présence de facteurs secondaires, influant sur le résultat de l'expérience, mais ne figurant pas parmi les conditions essentielles. Les conditions essentielles de l'expérience, déterminant en gros traits son développement, restent inchangées; les conditions secondaires varient d'une expérience à l'autre et introduisent des différences aléatoires dans les résultats.

Il est évident qu'il n'existe pas, dans la nature, d'effet physique libre de tout hasard. Quelque exactes et identiques que soient les conditions de l'expérience, il est impossible de faire de telle sorte que les résultats coïncident exactement et entièrement.

Tout phénomène déterministe est inévitablement accompagné d'écarts aléatoires. Néanmoins dans certaines applications pratiques, on peut négliger ces éléments aléatoires en remplaçant le phénomène réel par un schéma simplifié, son modèle, en supposant que dans les conditions données de l'expérience il se déroule d'une manière bien déterminée. Parmi la multitude des facteurs intervenant dans le phénomène examiné on sélectionne les plus importants, ceux qui sont fondamentaux et décisifs et on néglige tout simplement l'influence des autres facteurs. C'est la méthode usuelle

d'étude des phénomènes utilisée en physique, en mécanique et dans les applications techniques. Selon cette méthode, il s'agit avant tout de faire ressortir les conditions les plus importantes du problème et de trouver les paramètres qui en sont influencés; l'étape usuelle suivante est la mise en équations (on établit et on intègre par exemple, les équations différentielles décrivant le phénomène). On fait ainsi ressortir la loi essentielle régissant le phénomène étudié et permettant de prévoir le résultat de l'expérience pour des conditions données. Avec le progrès scientifique le nombre de facteurs dont on tient compte devient de plus en plus important, le phénomène est étudié plus en détail, la prédiction scientifique se fait plus précise.

Cependant pour résoudre un certain nombre de problèmes la méthode classique des sciences exactes ne peut être utilisée. Il existe des problèmes dans lesquels le résultat de l'expérience dépend d'un nombre de facteurs si important qu'il est impossible de les enregistrer tous et d'en tenir compte. C'est le cas des problèmes où les facteurs aléatoires secondaires intimement liés les uns aux autres jouent un rôle important, de plus, leur nombre est tellement grand et leur influence si compliquée que l'utilisation des méthodes classiques ne saurait être justifiée.

Considérons un exemple. Supposons que l'on tire sur une cible très importante, dont les dimensions sont grandes par rapport au domaine de dispersion des obus. Dans ce cas on peut négliger cette dispersion et faire simplement coïncider la trajectoire théorique avec le centre de la cible. Si au contraire (comme ceci a généralement lieu dans la pratique) le domaine de dispersion est supérieur aux dimensions de la cible, certains obus, sous l'effet des facteurs aléatoires, n'atteindront pas la cible. On peut alors se demander: quel est le pourcentage moyen des obus atteignant la cible? Quel est le nombre des obus qu'il y a lieu de tirer pour détruire la cible avec assez de sûreté? Quelles sont les mesures à prendre pour diminuer la dépense des obus?, etc.

Les méthodes des sciences exactes ne permettent pas de répondre à ces questions. Pour y répondre on ne peut pas tout simplement négliger le hasard, il faut étudier les lois régissant la dispersion aléatoire des trajectoires, trouver les causes et le caractère de la dispersion.

Soit un autre exemple. Considérons la recette journalière d'une station d'essence. La recette varie d'une manière aléatoire d'un jour à l'autre et ceci d'une manière imprévisible. En effet, elle dépend du nombre de voitures et de camions arrivés à la station pour faire leur plein d'essence, de la quantité d'essence que chacun d'eux prendra, du nombre de voitures passant devant la station sans s'y arrêter à cause d'une queue trop grande. Néanmoins, malgré le hasard il y a lieu de répondre à un grand nombre de questions. Par exemple,

quel doit être le nombre de pompes à la station? Quelle doit être la réserve de carburant de chaque type? Comment augmenter la rentabilité de la station? Et ainsi de suite.

Ces questions sont intimement liées à la nature aléatoire des effets, à l'irrégularité des arrivées, au caractère aléatoire de chaque commande (type de carburant, sa quantité), au hasard dans la formation et la dispersion des files d'attente, etc. Pour répondre à ces questions il est indispensable d'étudier le travail des stations d'essence non pas sous son aspect déterministe, mais compte tenu du hasard.

Soit encore un exemple. Un dispositif technique, par exemple, un système de commande automatique, est appelé à résoudre un certain problème, en présence de l'action continue de bruits aléatoires. Les bruits font que le système résout le problème posé avec une certaine erreur dépassant parfois la valeur tolérée. Il s'agit de déterminer la fréquence d'apparition des erreurs importantes et de prévoir les mesures à prendre pour pratiquement exclure leur éventualité.

La réponse à ces questions exige que soient étudiés la nature et la structure des perturbations aléatoires agissant sur le système, la réponse du système à ces perturbations, le rôle des paramètres constructifs du système dans la forme de cette réponse.

Pour résoudre ces problèmes, très nombreux en physique et en technique (tout comme dans d'autres domaines) il y a lieu d'étudier non seulement les lois principales déterminant le phénomène dans les lignes générales, mais également de procéder à l'analyse des perturbations et des déformations aléatoires liées à la présence de facteurs secondaires, introduisant une certaine indétermination dans le résultat de l'expérience pour des conditions données.

Quelles sont les méthodes permettant d'étudier les phénomènes aléatoires?

Du point de vue purement théorique, les facteurs, que nous avons conventionnellement appelés aléatoires, ne diffèrent en principe en rien des autres facteurs, que nous avons appelés essentiels. Théoriquement on peut augmenter indéfiniment la précision de la solution de chacun des problèmes, tenant compte d'un nombre de plus en plus grand de facteurs, des plus importants aux plus insignifiants. Néanmoins, une telle étude détaillée et scrupuleuse de l'influence de tous les facteurs, dont dépend l'effet, donnerait une solution trop compliquée, pratiquement irréalisable, sans augmenter notre connaissance de l'objet.

Il est évident qu'il doit y avoir une différence de principe dans les méthodes permettant de tenir compte des facteurs essentiels, déterminant le caractère principal du phénomène, et celles qui tiennent compte des facteurs secondaires dont l'influence sur le cours du phénomène se manifeste par une erreur ou une perturbation.

Le hasard, la complexité, la multitude des causes intervenant dans des phénomènes aléatoires rendent nécessaires des méthodes spéciales pour leur étude.

Ces méthodes sont élaborées dans la théorie des probabilités, qui a pour objet d'étude les lois régissant les phénomènes aléatoires.

L'expérience montre que lorsqu'on observe un grand nombre de phénomènes aléatoires, on y décèle généralement des lois tout à fait déterminées, en quelque sorte stables.

Voyons quelques exemples.

On jette de nombreuses fois une pièce de monnaie; la fréquence d'apparition de face pour un grand nombre d'expériences se stabilise peu à peu, tendant vers un nombre déterminé qui est dans ce cas égal à $^1/_2$. Cette même stabilité des fréquences apparaît également

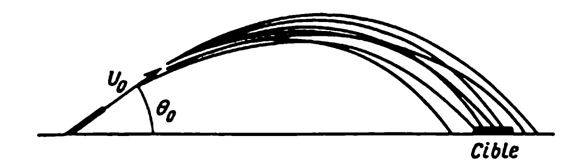


Fig. 1.1.2

lorsqu'une expérience quelconque dont l'issue est aléatoire, indéterminée à l'avance, est répétée de nombreuses fois. Ainsi, par exemple, pour un grand nombre de coups tirés, la fréquence de coups atteignant le but se stabilise, tendant vers un certain nombre constant.

Un récipient contient un certain volume de gaz composé par un grand nombre de molécules. Chaque molécule s'entrechoque de nombreuses fois avec les autres, change de nombreuses fois de vitesse et de direction; donc sa trajectoire est aléatoire. La pression du gaz sur les parois du récipient est due à l'ensemble des chocs des molécules contre les parois. Il pourrait sembler que si la trajectoire de chaque molécule est aléatoire, la pression sur les parois du récipient devrait également changer d'une manière aléatoire et incontrôlable, mais il n'en est rien. Si le nombre de molécules est suffisamment important, la pression du gaz ne dépend pratiquement pas des trajectoires des différentes molécules et suit une loi déterminée et très simple. Les particularités aléatoires propres au mouvement de chaque molécule se trouvent compensées dans leur ensemble de sorte que, malgré la complexité de chaque phénomène aléatoire particulier, on obtient une loi très simple, vraie pour leur ensemble. Notons que c'est justement la multitude des phénomènes aléatoires qui donne lieu à une loi déterminée; lorsque le nombre de molécules est restreint, on voit apparaître des écarts aléatoires, appelés fluctuations.

Des coups sont tirés les uns après les autres sur une cible; on étudie la répartition des points d'atteinte. Lorsque le nombre de coups est faible, les points d'atteinte sont répartis d'une manière désordonnée, sans qu'aucune loi se manifeste. Quand le nombre de coups augmente, on voit apparaître une certaine loi dans la disposition des points d'atteinte; cette loi est d'autant plus nette que le nombre de coups est important. La disposition des points d'atteinte est à peu près symétrique par rapport à un certain point central: dans la zone centrale les trous sont bien plus denses que sur les bords; la densité des trous décroît suivant une loi déterminée (cette loi est appelée loi normale ou loi de Gauss, et nous l'étudierons en détail dans notre cours).

Des lois spécifiques, dites *statistiques*, sont observées chaque fois que nous avons affaire à des effets aléatoires se produisant en grand nombre. Les lois apparaissant dans l'ensemble de phénomènes sont pratiquement indépendantes des particularités individuelles des composantes. Ces particularités se compensent dans la masse et le résultat moyen n'est plus pratiquement aléatoire. C'est cette stabilité, confirmée par l'expérience, qui légitime l'utilisation des méthodes probabilistes (statistiques) d'étude. Les méthodes de la théorie des probabilités sont de par leur nature adaptées à l'étude des phénomènes aléatoires dans leur ensemble, sans donner aucune idée sur le résultat d'une quelconque expérience aléatoire.

Plus grand est le nombre d'expériences aléatoires identiques étudiées dans un problème, plus nettes sont les lois inhérentes au phénomène, plus grandes sont la précision et la certitude des prévisions.

Dans tous les cas où l'on applique les méthodes probabilistes on tend à éviter l'étude trop compliquée (et parfois pratiquement impossible) d'un effet isolé, conditionné par un très grand nombre de facteurs, et on entreprend l'étude des lois régissant le phénomène aléatoire dans son ensemble. L'étude de ces lois permet non seulement de faire des prévisions scientifiques dans un domaine très particulier de phénomènes aléatoires, mais également de diriger le cours des processus aléatoires, de les contrôler, de limiter le champ d'action du hasard, de réduire son influence pratique.

La méthode probabiliste ou statistique d'étude ne s'oppose pas à la méthode classique, mais la complète, permettant d'analyser plus en détail le phénomène, compte tenu du hasard.

Le développement actuel des sciences naturelles et techniques est caractérisé par l'utilisation globale et fructueuse des méthodes statistiques. C'est tout naturel, car l'étude détaillée d'un phénomène quelconque ne peut se borner à celle des lois essentielles, il arrive un moment où il y a lieu d'analyser les écarts éventuels. Des branches entières de la physique moderne (en particulier la physique nucléaire) sont basées sur les méthodes de la théorie des probabilités. Les métho-

des probabilistes sont de plus en plus courantes en électrotechnique et radiotechnique modernes, en météorologie et astronomie, dans la théorie de commande automatique et dans la programmation, dans les problèmes économiques, sociaux, militaires, etc.

Les lois mathématiques de la théorie des probabilités reflètent les lois statistiques réelles. Les méthodes mathématiques efficaces et rigoureuses, mises en œuvre par la théorie des probabilités pour l'étude des phénomènes aléatoires, en font une branche mathématique aussi logique et précise que les autres sciences mathématiques.

1.2. Bref aperçu historique

La théorie des probabilités, comme beaucoup d'autres sciences mathématiques, est née des besoins pratiques.

C'est au XVII[e] siècle que commence l'étude systématique des problèmes liés aux phénomènes aléatoires et où l'on voit apparaître l'appareil mathématique correspondant. Au début du XVII[e] siècle l'éminent physicien Galilée essaie déjà d'étudier les erreurs des mesures physiques, considérant celles-ci comme aléatoires et estimant leur probabilité. A cette époque apparaît également la théorie des assurances, basée sur l'analyse des lois régissant des phénomènes aléatoires tels que la morbidité, la mortalité, les accidents, etc. La nécessité d'un appareil mathématique apte à analyser les phénomènes aléatoires se fait sentir dans toutes les sciences où le traitement et la généralisation d'un grand nombre de données statistiques posaient des problèmes assez difficiles.

Cependant la théorie des probabilités en tant que science mathématique a été fondée non pas sur les problèmes pratiques mentionnés qui sont pourtant trop compliqués pour qu'on puisse en dégager les lois. Il a fallu tout d'abord étudier les manifestations du hasard plus simples. L'étude systématique a commencé par les jeux de hasard. Depuis des temps très anciens ces jeux ont été créés par des générations entières de telle sorte que l'issue de l'expérience ne dépende pas des conditions de l'expérience et soit purement aléatoire. Les jeux de hasard fournissent des modèles particulièrement simples et clairs de phénomènes aléatoires, permettant dans la forme la plus pure d'observer et d'étudier les lois spécifiques les régissant ; de plus, la possibilité de répéter de nombreuses fois une même expérience assure une vérification expérimentale de ces lois dans les conditions réelles des expériences répétées. Jusqu'à présent les exemples empruntés aux jeux de hasard et les problèmes analogues dits de « tirage exhaustif » sont utilisés dans tous les manuels traitant de la théorie des probabilités en tant que modèles simplifiés des phénomènes aléatoires.

La naissance d'une véritable théorie des probabilités au sens moderne du terme peut être située vers 1650, avec les travaux de

Fermat (1601-1665), Pascal (1623-1662) et Huygens (1629-1695) consacrés à la théorie des jeux de hasard. Dans ces travaux on voit se former peu à peu les notions importantes de probabilité et d'espérance mathématique; leurs propriétés essentielles ont été établies ainsi que les méthodes de leur calcul. Les méthodes probabilistes ont avant tout été utilisées dans les problèmes d'assurance. Dès la fin du XVII[e] siècle les assurances se trouvent basées sur des méthodes scientifiques. Plus tard la théorie des probabilités a pénétré dans d'autres domaines.

Les travaux de Jacob Bernoulli (1654-1705) font un pas en avant dans le développement de la théorie des probabilités. On lui doit la démonstration de la loi fondamentale de la théorie des probabilités, appelée loi des grands nombres.

Avant Bernoulli on a remarqué une particularité des phénomènes aléatoires que l'on peut appeler « stabilité des fréquences pour un grand nombre d'expériences » et qui se manifeste en ce que pour un grand nombre d'expériences, ayant chacune une issue aléatoire, la fréquence relative d'apparition de chacune des issues tend à se stabiliser, convergeant vers un certain nombre, la probabilité de cette issue. Par exemple, si l'on jette un grand nombre de fois une pièce de monnaie, la fréquence relative d'apparition de face tend vers $1/2$; de même, lorsque l'on jette un grand nombre de fois un dé à six faces, la fréquence d'apparition de cinq points tend vers $1/6$, etc. Jacob Bernoulli était le premier à donner le fondement théorique de ce fait empirique. Le théorème de Jacob Bernoulli est la forme la plus simple de la loi des grands nombres, il établit une relation entre la probabilité d'un événement et la fréquence de son apparition; lorsque le nombre d'expériences est suffisamment important, on peut s'attendre, avec une certitude élevée, à ce que la fréquence s'approche aussi près que l'on veut de la probabilité.

Une autre étape importante dans le développement de la théorie des probabilités est liée au nom de Moivre (1667-1754). C'est lui qui a introduit et démontré, pour le cas le plus simple, une loi particulière, observée souvent dans les phénomènes aléatoires, appelée loi normale (ou loi de Gauss). Comme nous allons le voir, la loi normale joue un rôle particulièrement important dans les phénomènes aléatoires. Le groupe de théorèmes donnant un fondement à cette loi pour différentes conditions porte le nom général de « théorème central limite ».

Puis vinrent Laplace (1749-1827), Gauss (1777-1855), Poisson (1781-1840). Laplace a donné un exposé concis et systématique des fondements de la théorie des probabilités, démontré une des formes du théorème central limite (théorème de Moivre-Laplace) et développé de nombreuses applications remarquables de la théorie des probabilités aux questions pratiques, en particulier à l'analyse des erreurs d'observations et de mesures.

Gauss a donné un fondement encore plus sérieux à la loi normale, il a élaboré une méthode de traitement des données expérimentales appelée « méthode des moindres carrés ». Poisson a démontré une forme, plus générale que celle de Bernoulli, de la loi des grands nombres, et le premier a appliqué la théorie des probabilités aux problèmes du tir. Une des lois de répartition, jouant un rôle particulièrement important en théorie des probabilités et dans ses applications, porte le nom de Poisson.

Le XVIII[e] siècle et le début du XIX[e] siècle sont caractérisés par un développement intense de la théorie des probabilités ; à cette époque celle-ci devient à la mode. On l'utilise non seulement là où elle est applicable de droit, mais également là où elle n'est pas justifiée, comme, par exemple, dans les sciences morales. Un grand nombre de travaux virent le jour où la théorie des probabilités était appliquée au droit, à l'histoire, à la politique et même à la théologie. Dans tous ces travaux pseudo-scientifiques on perçoit une conception simplifiée, mécaniste des phénomènes sociaux. Des probabilités arbitraires se trouvaient à la base des raisonnements (par exemple, dans les questions de droit, on supposait que les gens mentent ou disent la vérité avec une probabilité égale pour tous) ; ceci posé, le problème social se résolvait comme un simple problème arithmétique. Il est évident que toutes ces tentatives étaient vouées à l'échec et ne pouvaient jouer un rôle positif dans le développement scientifique. Au contraire, le résultat indirect en a été que vers 1920-1930 la passion générale pour la théorie des probabilités s'est tournée en déception et scepticisme. Aux yeux du grand public elle est devenue une science douteuse, une sorte de distraction mathématique, non digne d'une étude sérieuse.

Il est remarquable que justement à cette époque apparaît célèbre école mathématique de Saint-Pétersbourg ayant donné à la théorie des probabilités son fondement logique et théorique et en ayant fait un outil efficace, exact et fiable de la connaissance. Dès la naissance de cette école le développement de la théorie des probabilités se trouve étroitement lié aux travaux des savants russes et soviétiques.

Parmi les savants de l'école mathématique de Saint-Pétersbourg il faut citer V. Bouniakovski (1804-1889) à qui l'on doit le premier cours de théorie des probabilités en langue russe, la terminologie russe de la théorie des probabilités et des travaux intéressants sur la statistique et la démographie.

A P. Tchébychev (1821-1894), éminent mathématicien russe, élève de V. Bouniakovski, on doit l'extension ultérieure et la généralisation de la loi des grands nombres ; de plus, il a introduit dans la théorie des probabilités la méthode puissante et féconde des moments.

A. Markov (1856-1922), élève de P. Tchébychev, a également enrichi la théorie des probabilités de découvertes et de méthodes fort

importantes. A. Markov a etendu le domaine d'application de la loi des grands nombres et du théorème central limite aux expériences indépendantes et dépendantes. Le mérite essentiel de A. Markov est d'avoir jeté les bases d'une branche nouvelle de la théorie des probabilités, à savoir de la théorie des processus aléatoires ou stochastiques.

Au nom de A. Liapounov (1857-1918), qui était aussi un élève de P. Tchébychev, est liée la première démonstration du théorème central limite pour des conditions très générales. Pour démontrer son théorème A. Liapounov a élaboré la méthode spéciale des fonctions caractéristiques, couramment utilisée de nos jours dans la théorie des probabilités.

Ces dernières décennies la théorie des probabilités est devenue l'une des sciences en plein essor, étroitement liée aux besoins pratiques et techniques. L'école probabiliste soviétique qui a hérité des traditions de l'école mathématique de Saint-Pétersbourg se trouve au premier plan de la science moderne.

Nous ne citerons ici que quelques-uns des éminents savants soviétiques dont les travaux ont joué un rôle décisif dans le développement de la théorie moderne des probabilités et de ses applications pratiques.

S. Bernstein (1880) a élaboré un système complet d'axiomes de la théorie des probabilités et a étendu le domaine d'application des théorèmes limites.

A. Khintchine (1894-1959) est bien connu pour ses travaux de généralisation et de renforcement de la loi des grands nombres, et surtout pour ses études des processus aléatoires stationnaires.

Un grand nombre de travaux fondamentaux dans différents domaines de la théorie des probabilités et de la statistique mathématique appartiennent à A. Kolmogorov (1903). C'est lui qui a correctement axiomatisé la théorie des probabilités en faisant ressortir ses liens avec une branche importante de l'analyse, à savoir la théorie métrique des fonctions. Les travaux de Kolmogorov sur la théorie des fonctions aléatoires (processus stochastiques) se trouvent actuellement à la base des études dans ce domaine et sont particulièrement importants. Ses travaux sur l'estimation de l'efficacité sont à la base d'une branche nouvelle de la théorie des tirs, devenue la science générale de l'efficacité des opérations militaires.

V Romanovski (1879-1954) et N. Smirnov sont connus pour leurs travaux de statistique mathématique, E. Sloutski (1880-1948) pour ses travaux sur la théorie des processus aléatoires, B. Gnédenko s'est consacré à l'étude des systèmes d'attente, E. Dynkine à l'étude des processus aléatoires markoviens, V. Pougatchev a beaucoup apporté à la théorie des processus aléatoires appliquée aux problèmes de commande automatique.

La théorie des probabilités est à l'heure actuelle en plein essor. Un intérêt particulier est porté aux questions relatives aux processus stochastiques. Des travaux importants dans ce domaine appartiennent à N. Wiener, W. Feller, J. Doob. Il y a lieu également de noter les travaux intéressants de R. Fisher, J. Neumann et H. Cramer sur la statistique mathématique et la théorie de commande automatique.

Ces dernières années on voit apparaître de nouvelles méthodes de la théorie des probabilités appliquée. Il s'agit en particulier de nouvelles disciplines, telle que la théorie de l'information et la théorie des systèmes d'attente. Nées des besoins pratiques, ces nouvelles branches acquièrent une importance théorique générale, leur domaine d'application ne cessant de s'étendre.

CHAPITRE 2

NOTIONS FONDAMENTALES DE LA THÉORIE DES PROBABILITÉS

2.1. Evénement. Probabilité d'un événement

Toute science développant la théorie générale d'un ensemble de phénomènes fait usage d'un certain nombre de notions fondamentales. Par exemple, en géométrie ce sont les notions de point, de droite, de ligne; en mécanique les notions de force, de masse, de vitesse, d'accélération, etc. Il est évident que toutes ces notions ne peuvent être définies en toute rigueur, car définir une notion signifie la rapporter à d'autres, plus connues. On conçoit aisément que le processus de définition de ces notions les unes par les autres doit s'arrêter quelque part, ayant abouti aux notions primaires auxquelles reviennent toutes les autres et qui sont elles-mêmes non pas définies, mais seulement précisées.

De telles notions de base existent également en théorie des probabilités. Nous allons commencer par introduire la notion d'*événement*.

En théorie des probabilités on entend par événement tout fait pouvant être réalisé ou ne pas être réalisé par une expérience.

Citons quelques exemples d'événements:

A — l'apparition de face lorsque l'on jette une pièce de monnaie;

B — l'apparition de face trois fois de suite lorsqu'on jette une pièce de monnaie trois fois;

C — l'atteinte d'une cible lorsqu'on tire un coup;

D — le tirage d'un as dans un jeu de cartes;

E — la détection d'un objectif dans un cycle d'exploration d'une station radar;

F — la rupture d'un fil pendant une heure de fonctionnement d'un métier à tisser.

En examinant les événements mentionnés ci-dessus on voit que chacun d'eux est en une certaine mesure plus ou moins probable; pour certains de ces événements on peut dire immédiatement lequel est plus probable et lequel l'est moins. Par exemple, on voit tout de suite que l'événement A est plus probable que B et D. Il n'en est pas de même pour les événements C, E et F, à cet effet il y aurait lieu de préciser les conditions de l'expérience. Il est évident que pour comparer quantitativement entre eux les événements suivant

leur possibilité de réalisation, il faut associer à chaque événement un certain nombre qui est d'autant plus grand que la possibilité de réalisation est plus élevée et que nous appellerons *probabilité de l'événement*.

Nous avons ainsi introduit la seconde notion fondamentale de la théorie des probabilités. La probabilité d'un événement est la mesure numérique de la possibilité effective de cet événement.

Notons que dès l'introduction de la notion de probabilité d'un événement, celle-ci est douée d'un sens pratique ; en effet, en se basant sur l'expérience on considère qu'un événement est plus probable s'il se produit plus souvent et moins probable s'il est plus rare et peu probable s'il n'apparaît presque jamais. Ainsi, la notion de probabilité d un événement est liée à la notion pratique de *fréquence* d'un événement.

Pour comparer entre eux différents événements suivant leur éventualité, on doit disposer d'une unité de mesure. Pour unité de mesure il est tout naturel de prendre la probabilité d'un *événement certain*, c'est-à-dire d'un événement qui doit obligatoirement se produire dans l expérience. A titre d'exemple d'événement certain on peut citer le fait qu'un jet d'un dé n'amène pas plus de six points.

Si l on attribue à l'événement certain une probabilité égale à l'unité, tous les autres événements, probables mais non certains, seront caractérisés par des probabilités inférieures à l'unité.

L'événement contraire de l'événement certain est appelé *événement impossible*, c'est un événement qui ne peut pas se produire dans l'expérience en question. A titre d'exemple d'événement impossible on peut citer l'apparition de douze points lorsqu'on jette un dé. Il est naturel d'attribuer à l'événement impossible une probabilité égale à zéro.

Nous avons ainsi établi l'unité de mesure des probabilités, à savoir la probabilité de l'événement certain, et la gamme de variation des probabilités de tous les événements, les nombres de 0 à 1.

2.2. Calcul direct des probabilités

Pour toute une classe d'expériences, les probabilités des résultats éventuels peuvent facilement être estimées directement à partir des conditions de l'expérience. Pour cela il faut que les différents résultats des expériences soient symétriques et par conséquent équiprobables.

Considérons par exemple une expérience consistant à jeter un dé, c'est-à-dire un cube symétrique sur les faces duquel sont portés 1, 2, 3, 4, 5 et 6 points.

Vu la symétrie du dé, il y a lieu de considérer tous les six résultats de l'expérience équiprobables. C'est ce qui nous donne le droit de supposer que lorsqu'on jette un dé de nombreuses fois toutes les

six faces sortiront un même nombre de fois. Pour un dé bien équilibré cette hypothèse est vérifiée par l'expérience; lorsqu'on jette un dé un grand nombre de fois chacune de ses faces apparaît à peu près une fois sur six, l'écart par rapport à la valeur $^1/_6$ est d'autant plus petit que le nombre d'expériences est important. Comme la probabilité d'un événement certain est égale à l'unité, il est tout naturel d'attribuer la probabilité $^1/_6$ à l'apparition de chacune des faces. Ce nombre caractérise les propriétés effectives du phénomène aléatoire étudié, à savoir la symétrie des six résultats possibles de l'expérience.

Ainsi, pour toutes les expériences dont les résultats possibles sont symétriques et équiprobables, on peut appliquer une méthode analogue appelée *calcul direct des probabilités*.

En général, on ne peut observer une symétrie dans les résultats éventuels des expériences que si ces dernières sont artificielles, comme dans le cas des jeux de hasard. La théorie des probabilités ayant été développée justement sur la base des jeux de hasard, la méthode de calcul direct des probabilités apparue en même temps que la théorie des phénomènes aléatoires, a été durant très longtemps considérée comme fondamentale, c'est sur elle que se basait la théorie « classique » des probabilités. Les expériences qui ne possédaient pas la symétrie des résultats possibles ont été rapportées artificiellement au schéma classique.

Bien que le domaine d'application pratique de ce schéma soit limité, il présente un certain intérêt, car les expériences à résultats éventuels symétriques et les événements liés à ces expériences, permettent de mieux comprendre les propriétés essentielles des probabilités. Nous allons tout d'abord étudier les événements de ce genre auxquels peut s'appliquer le schéma de calcul direct des probabilités.

A cet effet nous introduisons certaines notions auxiliaires.

1. Système complet d'événements.

On dit que plusieurs événements forment un *système complet d'événements* si, dans une expérience, un événement et un seul du système est réalisé à la fois.

A titre d'exemple d'événements formant un système complet on peut citer:

1) l'apparition de pile ou de face lorsqu'on jette une pièce de monnaie;

2) un coup porté ou un coup raté lors d'un tir;

3) l'apparition de 1, 2, 3, 4, 5, 6 points dans le jet d'un dé;

4) le tirage d'une boule blanche ou d'une boule noire lorsque l'on tire une boule d'une urne contenant par exemple 2 boules blanches et 3 boules noires;

5) zéro, une, deux, trois ou plus d'erreurs d'impression dans une page d'épreuve;

6) de deux coups tirés au moins un coup atteignant le but ou un coup raté.

2. Evénements incompatibles.

Des événements sont dits *incompatibles* dans une expérience si aucuns deux d'entre eux ne peuvent apparaître simultanément.

A titre d'exemple d'événements incompatibles on peut citer:

1) l'apparition simultanée de pile et de face lorsqu'on jette une pièce de monnaie;

2) atteindre et rater simultanément une cible par un seul coup;

3) l'apparition de 1, 2, 3 points simultanément lors du jet d'un dé ;

4) exactement une, deux, trois pannes simultanées d'un dispositif technique pour 10 heures de fonctionnement.

3. Evénements équiprobables.

Plusieurs événements dans une expérience sont dits *équiprobables* si la symétrie donne lieu de supposer qu'aucun de ces événements n'est plus probable qu'un autre.

A titre d'exemple d'événements équiprobables on peut citer:

1) l'apparition de pile et de face lorsqu'on jette une pièce de monnaie;

2) l'apparition de 1, 2, 3, 4, 5, 6 points quand on jette un dé;

3) le tirage d'un carreau, d'un cœur, d'un trèfle ou d'un pique dans un jeu de cartes;

4) le tirage d'une boule portant le n° 1, 2, 3, . . . lorsqu'on tire des boules d'une urne contenant 10 boules numérotées.

Il existe des événements doués des trois propriétés mentionnées; ils forment un système complet, sont incompatibles et équiprobables; par exemple, l'apparition de pile ou de face lorsqu'on jette une pièce de monnaie; l'apparition de 1, 2, 3, 4, 5, 6 points lorsqu'on jette un dé. Les événements formant un tel groupe sont dits des *cas*.

Pour une expérience à résultats éventuels symétriques, les cas forment un système exhaustif de résultats équiprobables et mutuellement incompatibles. On dit qu'une telle expérience se réduit à un « système de cas » ou encore, à un « tirage exhaustif ».

Le plus souvent on rencontre des systèmes de cas dans les expériences artificielles, dans lesquelles l'équiprobabilité des résultats est assurée d'avance (comme par exemple, dans les jeux de hasard). Dans de telles expériences on peut directement calculer la probabilité en estimant le nombre des *cas favorables* parmi tous les *cas possibles*.

Un cas est dit *favorable* à un événement si son apparition entraîne la réalisation de l'événement en question.

Par exemple, lorsque l'on jette un dé, six cas sont possibles: apparition de 1, 2, 3, 4, 5, 6 points. Si l'événement A consiste en l'apparition d'un nombre pair de points, les cas de 2, 4, 6 points sont favorables, les trois autres étant défavorables.

Si une expérience se réduit à un système de cas, la probabilité de l'événement A dans cette expérience peut être estimée comme la

proportion des cas favorables. La probabilité $P(A)$ de l'événement A se calcule comme le rapport du nombre m de cas favorables au nombre n de tous les cas possibles:

$$P(A)=\frac{m}{n}. \tag{2.2.1}$$

Le nombre de tous les cas favorables se trouvant toujours entre 0 et n (0 pour l'événement impossible et n pour l'événement certain), la probabilité de l'événement calculée par la formule (2.2.1) est toujours une fraction rationnelle propre:

$$0 \leqslant P(A) \leqslant 1. \tag{2.2.2}$$

La formule (2.2.1) dite « formule classique » de calcul des probabilités, a longtemps servi à définir la probabilité. A l'heure actuelle on se base, généralement, sur d'autres principes, liant directement la notion de probabilité à la notion empirique de fréquence relative; quant à la formule (2.2.1), elle s'est conservée comme une formule de calcul direct des probabilités quand l'expérience se réduit à un système de cas, c'est-à-dire lorsque les issues possibles de l'expérience sont équiprobables.

Exemple 1. Une urne contient 2 boules blanches et 3 noires. On tire au hasard une boule de l'urne. Trouver la probabilité de tirer une boule blanche.

Solution. Nous allons désigner par A l'événement correspondant au tirage d'une boule blanche. Le nombre des cas possibles est $n = 5$; le nombre des cas favorables à l'événement A est $m = 2$. Par conséquent:

$$P(A)=\frac{2}{5}.$$

Exemple 2. Une urne contient a boules blanches et b boules noires. On tire deux boules de l'urne. Trouver la probabilité de tirer deux boules blanches.

Solution. Désignons par B l'événement correspondant au tirage de deux boules blanches. Calculons le nombre total des cas possibles n et le nombre m des cas favorables à l'événement B:

$$n=C_{a+b}^{2};$$

$$m=C_{a}^{2}\ ^{*}),$$

par conséquent.

$$P(B)=\frac{C_a^2}{C_{a+b}^2}.$$

Exemple 3. Dans un lot de N pièces. M sont défectueuses. On prend dans ce lot n pièces au hasard. Trouver la probabilité pour que, parmi n pièces, on trouve exactement m défectueuses.

---

*) Le signe C_k^l désigne le nombre de combinaisons de k éléments pris l à l.

Solution. Le nombre total des cas est de toute évidence égal à C_N^n; le nombre des cas favorables est $C_M^m C_{N-M}^{n-m}$, d'où la probabilité de l'événement qui nous intéresse

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

2.3. Fréquence relative, ou probabilité statistique d'un événement

La formule (2.2.1) pour le calcul direct des probabilités n'est applicable que si l'expérience pouvant réaliser l'événement qui nous intéresse a des issues possibles symétriques (se réduit à un système de cas). Il est clair que toutes les expériences ne peuvent pas se réduire à un système de cas, et il existe une classe importante d'événements dont les probabilités ne peuvent pas être calculées par la formule (2.2.1). Considérons, par exemple, un dé mal équilibré. L'apparition de telle ou telle face n'est plus alors caractérisée par la probabilité $^1/_6$; on comprend intuitivement que pour ce dé mal équilibré l'apparition de la face défectueuse possède une certaine probabilité, indiquant la fréquence moyenne d'apparitions de cette face lorsque le dé est jeté de nombreuses fois. De même, les probabilités d'événements tels que l'atteinte de la cible lors d'un coup, le défaut d'un tube électronique durant une heure de fonctionnement ou l'apparition du signal « occupé » sur un appel téléphonique ne peuvent pas non plus être calculées à l'aide de la formule (2.2.1), car les expériences correspondantes ne se réduisent pas à un système de cas. Cependant il est évident que chacun des événements énumérés possède une certaine possibilité effective, qui en principe peut être calculée numériquement et qui, si l'on multiplie ces expériences, sera donnée par la fréquence relative des événements correspondants. Nous allons supposer alors que chaque événement lié à un grand nombre d'expériences homogènes, se réduisant ou non à un système de cas, a une certaine probabilité se trouvant entre zéro et l'unité. Pour les événements se réduisant à un système de cas cette probabilité peut être calculée directement à l'aide de la formule (2.2.1). Pour les événements qui ne se réduisent pas à un système de cas on utilise d'autres méthodes. Toutes ces méthodes ont leurs racines dans l'expérience et pour s'en faire une idée il y a lieu de préciser la notion de fréquence d'un événement et le caractère particulier de la relation existant entre la probabilité et la fréquence.

Soit une série de n expériences chacune desquelles peut réaliser ou ne pas réaliser un certain événement A. On appelle *fréquence* de l'événement A dans la série d'expériences en question le rapport du nombre d'expériences où l'événement A est apparu, au nombre total d'expériences effectuées.

La fréquence d'un événement est souvent appelée sa *probabilité statistique*, ou *empirique*, à la différence de la probabilité *mathématique* introduite ci-dessus.

Nous désignerons par $P^*(A)$ la fréquence (la probabilité statistique) de l'événement A. La fréquence d'un événement est calculée à partir des résultats des expériences par la formule suivante:

$$P^*(A) = \frac{m}{n}, \qquad (2.3.1)$$

où m est le nombre d'apparitions de l'événement A;
n le nombre total d'expériences.

Lorsque le nombre d'expériences n'est pas important la fréquence de l'événement a un caractère aléatoire et peut notablement changer d'un groupe d'expériences à l'autre. Par exemple, si l'on jette une pièce de monnaie 10 fois, il est fort probable que face apparaisse deux fois seulement (la fréquence d'apparition de face sera alors égale à 0,2); dix autres coups peuvent amener face huit fois (fréquence 0,8). Cependant lorsque le nombre d'expériences augmente la fréquence de l'événement perd son caractère aléatoire; les conditions aléatoires intervenant dans chaque expérience isolée se trouvant mutuellement compensées et la fréquence tend à se stabiliser, s'approchant d'une certaine grandeur constante. Par exemple, lorsqu'on jette une pièce de monnaie un grand nombre de fois, la fréquence d'apparition de face s'écartera peu de 1/2.

Cette « stabilité des fréquences », vérifiée maintes fois par les observations et confirmée par toute l'expérience humaine est une des lois les plus caractéristiques des phénomènes aléatoires. Cette loi a été formulée pour la première fois par J. Bernoulli dans le théorème portant son nom et représentant en fait une forme simple de la loi des grands nombres. J. Bernoulli a démontré que pour un grand nombre d'expériences homogènes indépendantes on peut affirmer avec certitude que la fréquence d'un événement différera aussi peu que l'on veut de sa probabilité dans une expérience isolée.

Le lien existant entre la fréquence d'un événement et sa probabilité est très profond. Ces deux notions sont pratiquement inséparables. En effet, lorsqu'on estime la probabilité d'un événement, on se base forcément sur la fréquence des événements analogues dans la pratique. En caractérisant la probabilité d'un événement par un nombre quelconque on ne peut attribuer à ce nombre une autre valeur réelle et une autre signification que la fréquence relative d'apparition de l'événement en question dans un grand nombre d'expériences. L'estimation numérique de l'éventualité d'un événement par la probabilité n'a de sens pratique que parce que les événements plus probables se produisent en moyenne plus souvent que les moins probables. Et si la pratique montre que, au fur et à mesure de l'augmentation du nombre d'expériences, la fréquence de l'événement

a tendance à se stabiliser, s'approchant rigoureusement d'un nombre constant, il est naturel d'admettre que ce nombre est la fréquence de l'événement.

Il est clair que l'on peut vérifier cette hypothèse seulement pour des événements dont les probabilités sont susceptibles d'être calculées directement, c'est-à-dire pour les événements se réduisant à un système de cas, car ce n'est qu'alors qu'il existe une méthode permettant de calculer la probabilité mathématique. De nombreuses expériences réalisées depuis la naissance du calcul des probabilités confirment effectivement cette hypothèse. Il est tout naturel d'admettre que pour les événements ne se réduisant pas à un système de cas, la même loi reste vraie et que le nombre constant vers lequel tend la fréquence d'un événement, lorsque le nombre d'expériences augmente, n'est rien d'autre que la probabilité de l'événement. On peut alors prendre la fréquence d'un événement, calculée pour un nombre suffisamment grand d'expériences, pour la valeur approchée de la probabilité. C'est ce qu'on fait en pratique, en déterminant à partir de l'expérience les probabilités des événements ne se réduisant pas à un système de cas.

Il y a lieu de noter que la tendance de la fréquence relative vers la probabilité avec augmentation du nombre d'expériences diffère quelque peu d'une tendance vers une limite au sens mathématique du terme.

Lorsqu'en mathématiques on dit que la variable X_n tend vers une limite constante a, lorsque n augmente, ceci signifie que la différence $|X_n - a|$ devient inférieure à un nombre positif quelconque ε pour tous les n supérieurs à un certain nombre suffisamment grand.

Pour ce qui est de la fréquence relative d'un événement et de sa probabilité, une telle affirmation catégorique n'est pas possible. En effet, il n'est physiquement pas impossible que même pour un grand nombre d'expériences la fréquence d'un événement s'écartera notablement de sa probabilité ; mais un tel écart important est assez peu probable, et d'autant moins que le nombre d'expériences est grand. Par exemple, lorsque l'on jette une pièce de monnaie dix fois, il est physiquement possible (bien que peu probable) d'amener toutes les dix fois face et donc d'obtenir la fréquence d'apparition de face égale à 1 ; lorsque l'on jette cette pièce 1000 fois, cet événement reste encore physiquement possible, mais il est tellement peu probable que l'on peut aisément le considérer comme pratiquement irréalisable. Ainsi, lorsque le nombre d'expériences augmente, la fréquence tend vers la probabilité, bien que non pas avec une certitude cent pour cent, mais avec une probabilité assez élevée pour que, avec un nombre suffisamment grand d'expériences, on puisse la considérer comme une certitude.

En théorie des probabilités la manière dont très souvent les grandeurs tendent les unes vers les autres porte un caractère particu-

lier et pour le décrire on utilise la notion *de convergence en probabilité.*

On dit que la grandeur X_n *tend* (ou converge) *en probabilité* vers la grandeur a si pour un ε aussi petit que l'on veut la probabilité de l'inégalité $|X_n - a| < \varepsilon$ tend vers l'unité lorsque n augmente indéfiniment.

En utilisant ce terme, on peut dire que lorsque le nombre d'expériences augmente, la fréquence d'un événement ne *tend* pas vers sa probabilité, mais *converge* en probabilité vers ce nombre.

Cette propriété de convergence de la fréquence en probabilité, exposée ici sans fondements mathématiques suffisants, sur des bases pratiques et le bon sens, est l'objet du théorème de Bernoulli qui sera démontré dans le chapitre 13.

Ainsi, en introduisant la notion de fréquence d'un événement et en utilisant la relation existant entre la fréquence et la probabilité, on peut attribuer une certaine probabilité comprise entre zéro et l'unité, non seulement aux événements se réduisant à un système de cas, mais également à ceux qui ne s'y réduisent pas; dans ce dernier cas la probabilité d'un événement peut approximativement être déterminée par la fréquence de l'événement pour un grand nombre d'expériences.

Nous verrons ultérieurement que pour trouver la probabilité d'un événement ne se réduisant pas à un système de cas, il n'est pas toujours indispensable de déterminer la fréquence directement à partir de l'expérience. La théorie des probabilités dispose de nombreuses méthodes permettant de trouver indirectement les probabilités en fonction des probabilités d'autres événements liés à ceux-ci. Ce sont ces méthodes indirectes qui sont l'objet de la théorie des probabilités. Cependant même ces méthodes indirectes font, en fin de compte, appel à l'expérience. La fiabilité et la valeur effective des calculs pratiques menés à la base de la théorie des probabilités sont déterminées par la qualité et la quantité des données expérimentales utilisées.

De plus, pour les applications pratiques des méthodes probabilistes, il faut être sûr que le phénomène aléatoire étudié est réellement un effet de masse pour lequel, du moins dans un certain intervalle de temps, la fréquence est stable. Ce n'est que dans ce cas qu'on peut parler des probabilités des événements en entendant par là non pas des fictions mathématiques mais les caractéristiques réelles des phénomènes aléatoires.

Par exemple, l'affirmation « le calcul mené sur ordinateur durant 10 heures avec une probabilité de 0,9 ne contient pas d'erreurs » a un sens bien déterminé, car de tels calculs sont des opérations courantes se répétant de nombreuses fois dans des conditions analogues.

Au contraire, l'affirmation « un problème scientifique est résolu correctement avec une probabilité de 0,9 » n'a aucun sens, car la solution d'un problème scientifique est une opération individuelle et il serait une erreur d'estimer par les méthodes probabilistes la vraisemblance des fondements scientifiques.

2.4. Variable aléatoire

Une des notions fondamentales de la théorie des probabilités est celle de variable aléatoire.

On appelle *variable aléatoire* une grandeur qui peut dans l'expérience prendre l'une quelconque des valeurs possibles, inconnue d'avance.

Exemples de variables aléatoires:

1) le nombre de coups réussis parmi trois coups tirés;
2) le nombre d'appels journalier d'un central téléphonique;
3) la fréquence de face dans une partie de 10 coups.

Dans les trois exemples mentionnés les variables aléatoires peuvent prendre l'une des valeurs qu'on peut énumérer d'avance.

Ainsi, dans l'exemple 1) ces valeurs sont:

$$0,\ 1,\ 2,\ 3;$$

dans l'exemple 2) ce sont:

$$1,\ 2,\ 3,\ 4,\ \ldots$$

et dans l'exemple 3):

$$0;\ 0{,}1;\ 0{,}2;\ \ldots;\ 1{,}0.$$

Des variables aléatoires prenant des valeurs discrètes, pouvant être énumérées d'avance, sont dites *discrètes* ou *discontinues*.

Il existe également des variables aléatoires d'un autre type, par exemple:

1) l'abscisse d'un point se déplaçant le long de l'axe des abscisses d'une façon aléatoire;
2) l'erreur de pesée d'un corps sur une balance automatique;
3) la vitesse d'un avion à une altitude donnée;
4) le poids d'une graine de blé prise au hasard.

Les valeurs éventuelles de ces variables aléatoires ne sont pas séparées les unes des autres; elles remplissent d'une manière continue un certain intervalle qui parfois a des limites très nettes, le plus souvent celles-ci sont indéterminées.

Des variables aléatoires dont les valeurs possibles remplissent d'une manière continue un certain intervalle sont dites *continues*.

La notion de variable aléatoire joue un rôle très important en théorie des probabilités. Si la théorie classique des probabilités opérait principalement avec des événements, au contraire la théorie

moderne préfère là, où il est possible, d'utiliser des variables aléatoires.

Nous allons donner quelques exemples typiques pour la théorie des probabilités où l'on passe d'un événement à une variable aléatoire.

Soit une expérience dans laquelle l'événement A peut apparaître ou ne pas apparaître. Au lieu de l'événement A on peut envisager la variable aléatoire X, égale à 1 si l'événement A a lieu, et à 0 dans le cas contraire. Cette variable aléatoire est de toute évidence discrète, car elle prend deux valeurs possibles: 0 et 1. Elle est appelée *variable indicatrice* de l'événement A. Dans la pratique, il est souvent plus commode d'utiliser au lieu des événements les variables indicatrices qui leur sont associées. Par exemple, si l'on effectue une série d'expériences dans chacune desquelles on peut voir apparaître l'événement A, le nombre total d'apparitions de l'événement est égal à la somme des variables indicatrices de l'événement A dans toutes les expériences.

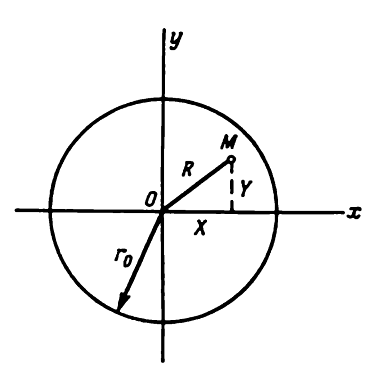


Fig. 2.4.1

D'un autre côté, très souvent, pour calculer la probabilité d'un événement il est commode d'associer à cet événement une variable aléatoire continue (ou un système de variables continues).

Supposons que l'on mesure les coordonnées d'un certain objet O pour construire le point M le représentant sur un dessin panoramique d'une région. Nous nous intéressons à l'événement A consistant en ce que l'erreur R dans la position du point M ne sera pas supérieure à une certaine valeur r_0 (fig. 2.4.1). Désignons par X, Y les erreurs aléatoires de mesure des coordonnées de l'objet. Il est clair que l'événement A équivaut à ce que le point aléatoire M de coordonnées X, Y se trouve à l'intérieur d'un cercle de rayon r_0 de centre O. En d'autres termes, pour que l'événement A ait lieu les variables aléatoires X et Y doivent satisfaire à l'inégalité:

$$X^2+Y^2<r_0^2. \qquad (2.4.1)$$

La probabilité de l'événement A n'est rien d'autre que la probabilité de vérification de l'inégalité (2.4.1). Cette probabilité peut être déterminée si l'on connaît les propriétés des variables aléatoires X, Y.

Ce lien étroit entre les événements et les variables aléatoires est caractéristique pour la théorie moderne des probabilités, qui chaque fois que c'est possible passe d'un système d'événements à un

système de variables aléatoires. Cette dernière méthode est bien plus souple et plus universelle que la première et convient mieux à la résolution des problèmes relatifs aux phénomènes aléatoires.

2.5. Evénements presque impossibles et événements presque certains. Principe de certitude pratique

Dans le § 2.2 nous avons introduit les notions d'événement impossible et d'événement certain. La probabilité de l'événement impossible, égale à zéro, et la probabilité de l'événement certain, égale à l'unité, sont deux limites extrêmes de l'échelle des probabilités.

Dans la pratique on a souvent affaire non pas à des événements impossibles et certains, mais à des événements « presque impossibles » et « presque certains ».

On appelle *événement presque impossible* un événement dont la probabilité n'est pas exactement égale à zéro, mais en est très voisine.

Soit, par exemple, l'expérience suivante: 32 lettres d'un alphabet *), écrites chacune sur une carte isolée, sont mélangées dans une urne; on tire une carte au hasard, on inscrit la lettre qu'elle porte et on remet la carte dans l'urne que l'on secoue pour bien mélanger les cartes. On répète 38 fois cette expérience. Considérons l'événement A consistant en ce que, après ces 38 tirages non exhaustifs, on verra se composer la première ligne de « Tartuffe » de Molière:

« Allons, Flipote, allons, que d'eux je me délivre ».

Du point de vue logique cet événement n'est pas impossible: on peut calculer sa probabilité qui est égale à $\left(\frac{1}{32}\right)^{38}$; mais cette probabilite est tellement petite que l'on peut considérer l'événement A comme presque impossible.

On appelle *événement presque certain* un événement dont la probabilité n'est pas exactement égale à l'unité, mais en est très voisine.

Si un événement quelconque $\bar{A}$ dans une expérience est presque impossible, l'événement contraire A, équivalent à la non-réalisation de A, sera presque certain. Ainsi, du point de vue de la théorie des probabilités, on peut indifféremment parler d'événements presque impossibles ou d'événements presque certains, car ils accompagnent toujours les uns les autres.

Les événements presque impossibles et presque certains jouent un rôle très important dans la théorie des probabilités; toutes les applications pratiques sont basées sur ces notions.

En effet, si l'on sait que la probabilité d'un événement dans une expérience est égale à 0,3, cela ne permet pas de prévoir le résultat

---

*) On suppose qu'il s'agit de l'alphabet français, plus les caractères é, è, à, ù, ç et l'apostrophe (la virgule).

de l'expérience. Mais si la probabilité d'un événement dans une expérience est insignifiante ou au contraire très voisine de l'unité, on peut prévoir le résultat de l'expérience; dans le premier cas on ne peut pas s'attendre à l'apparition de l'événement A ; dans le second on a toutes les raisons de le voir se produire. Dans ce pronostic on se base sur le *principe de certitude pratique* qui peut être formulé comme suit.

Si la probabilité d'un événement A quelconque dans une expérience donnée E est suffisamment petite, on peut être presque certain que lorsque l'expérience E est effectuée une seule fois, l'événement A n'aura pas lieu.

En d'autres termes, si la probabilité de l'événement A dans une expérience donnée est très petite, on peut mener cette dernière comme si cet événement était impossible, c'est-à-dire sans tenir compte de sa possibilité de réalisation.

Dans l'expérience quotidienne on utilise toujours, bien qu'inconsciemment, le principe de certitude pratique. Par exemple, en partant en voyage par chemin de fer on ne tient pas compte de la possibilité d'une catastrophe, bien qu'une probabilité infime de cet événement existe.

Le principe de certitude pratique ne peut être démontré par des méthodes mathématiques; il se trouve confirmé par toute l'expérience humaine.

La question de savoir à quel point la probabilité d'un événement doit être petite pour que celui-ci puisse être considéré comme impossible sort du cadre de la théorie mathématique et, dans chaque cas, doit être résolue à partir de considérations pratiques conformément à l'importance du résultat désiré.

Par exemple, si la probabilité d'explosion intempestive d'une mine est égale à 0,01, on peut encore l'admettre et considérer la défectuosité qui en est cause comme un événement pratiquement impossible. Au contraire, si la probabilité de non-ouverture d'un parachute lors d'un saut est égale à 0,01, on ne peut admettre que le défaut est un événement pratiquement impossible et il y a lieu d'augmenter la fiabilité du mécanisme du parachute.

L'un des problèmes essentiels de la théorie des probabilités est la recherche des événements pratiquement impossibles (ou pratiquement certains) permettant de prévoir le résultat de l'expérience ainsi que la recherche des conditions pour lesquelles tel ou tel événement devient presque impossible (ou certain). En théorie des probabilités il existe un certain nombre de théorèmes, appelés théorèmes limites, établissant l'existence des événements presque impossibles (ou certains) lorsque le nombre d'expériences ou le nombre de variables aléatoires englobées par le problème augmente. A titre d'exemple on peut citer le théorème déjà mentionné de Bernoulli (une forme simple de la loi des grands nombres). Rappelons qu'en vertu du

théorème de Bernoülli, pour un grand nombre d'expériences, l'événement consistant en ce que la différence entre la fréquence d'un événement et sa probabilité soit infiniment petite devient presque certain.

En plus des événements presque impossibles (certains) permettant de prévoir avec certitude le résultat de l'expérience malgré le hasard, on envisage en théorie des probabilités des variables aléatoires d'un type particulier jouant un rôle important. Ces variables, bien qu'aléatoires, ont des variations tellement insignifiantes qu'on peut les considérer comme pratiquement non aléatoires, ou presque certaines. A titre d'exemple on peut citer la fréquence d'un événement pour un grand nombre d'expériences. Bien qu'aléatoire, pour un grand nombre d'expériences, ce nombre varie dans des limites étroites au voisinage de la probabilité de l'événement.

Des variables presque certaines permettent de prédire le résultat numérique de l'expérience, malgré la présence du hasard, et d'opérer avec ce résultat avec certitude, tout comme on le fait avec des données fournies par des méthodes des sciences exactes.

CHAPITRE 3

THÉORÈMES FONDAMENTAUX DE LA THÉORIE DES PROBABILITÉS

3.1. Rôle des théorèmes fondamentaux. Somme et produit d'événements

Dans le chapitre précédent nous avons étudié les méthodes directes de calcul des probabilités, à savoir la formule classique pour la probabilité d'un événement appartenant à un système complet d'événements (il s'agit du système de « cas » ou du schéma du tirage exhaustif) et la méthode de calcul approché de la probabilité d'après la fréquence relative d'un événement ne se réduisant pas à un système de cas. Cependant, ces méthodes directes ne jouent qu'un rôle médiocre dans la théorie des probabilités, car elles ne sont pas toujours commodes et sont parfois inutilisables. Même lorsque l'événement est tel que le schéma du tirage exhaustif est applicable, le calcul direct de la probabilité par la formule (2.2.1) s'avère parfois trop compliqué. Pour ce qui est des événements auxquels ce schéma n'est pas applicable, leurs probabilités sont très rarement calculées directement à l'aide des fréquences. Dans les applications pratiques il faut souvent déterminer les probabilités des événements qu'il est difficile de reproduire expérimentalement. Par exemple, s'il y a lieu de déterminer la probabilité d'abattre un avion dans un combat aérien, il est évident que le calcul de cette probabilité d'après la fréquence de réalisation est pratiquement impossible ; ceci non seulement parce que les expériences se seraient avérées trop compliquées et coûteuses, mais également parce qu'on entreprend les calculs de la probabilité précisément pour prédire le résultat de l'expérience pour des dispositifs techniques au stade du projet, pour choisir les paramètres constructifs les plus rationnels d'un dispositif projeté.

C'est pourquoi la plupart du temps on utilise des méthodes dites *indirectes*, permettant de trouver d'après les probabilités connues des événements déjà étudiés les probabilités d'autres événements, liés aux premiers. Toute la théorie des probabilités se réduit grossièrement à un système de méthodes indirectes permettant de réduire l'expérience au minimum.

En appliquant ces méthodes indirectes on utilise sous une forme ou une autre les *théorèmes fondamentaux* de la théorie des probabilités. Il s'agit du théorème d'addition des probabilités et du théorème de multiplication des probabilités. En toute rigueur, ces deux théorè-

mes peuvent être démontrés seulement pour le cas du tirage exhaustif; pour tous les autres cas ils doivent être adoptés comme des assertions ou des postulats.

Avant de formuler et de démontrer les théorèmes fondamentaux, nous introduisons des notions auxiliaires de *somme* et de *produit* d'événements.

Dans de nombreux domaines des sciences exactes on utilise des opérations symboliques qui reçoivent leur appellation par analogie avec les opérations arithmétiques similaires, dont elles possèdent un certain nombre de propriétés. Telles sont, par exemple, les opérations d'addition et de multiplication des vecteurs en mécanique, les opérations d'addition et de multiplication des matrices en algèbre, etc. Ces opérations, soumises à certaines lois, permettent non seulement de simplifier l'écriture, mais dans de nombreux cas d'alléger la structure logique des déductions scientifiques. L'introduction des opérations symboliques sur les événements est également de grande utilité en théorie des probabilités.

On appelle *somme de deux événements* A et B et on note $A + B$ l'événement C consistant en ce que soit réalisé ou bien l'événement A, ou bien l'événement B, ou les deux ensemble.

Par exemple, si l'événement A est l'atteinte d'une cible lors du premier coup, l'événement B l'atteinte de la cible lors du second coup, alors l'événement $C = A + B$ traduit le fait que la cible est atteinte, fût-ce avec le premier, le second ou les deux coups à la fois.

Si les événements A et B sont incompatibles, ils ne peuvent de toute évidence se produire simultanément et la somme $A + B$ exprime la réalisation soit de l'événement A, soit de l'événement B. Si par exemple A est « tirage d'un cœur », et B « tirage d'un carreau», l'événement $C = A + B$ est « tirage d'une carte de couleur rouge, cœur ou carreau ».

Résumons: on appelle *somme de deux événements* A et B l'événement C correspondant à la réalisation de l'un des événements A ou B.

La *somme de plusieurs événements* est l'événement correspondant à la réalisation au moins d'un des événements en question.

Par exemple l'expérience consiste à tirer cinq coups sur une cible. Enumérons les événements qui peuvent se produire:

A_0 — aucun coup n'atteint le but;
A_1 — exactement un coup atteint le but;
A_2 — exactement deux coups atteignent le but;
A_3 — exactement trois coups atteignent le but;
A_4 — exactement quatre coups atteignent le but;
A_5 — tous les cinq coups atteignent le but.

L'événement

$$A = A_0 + A_1 + A_2$$

se lit « deux coups au plus atteignent le but » et l'événement

$$B = A_3 + A_4 + A_5$$

se lit « trois coups au moins atteignent le but ».

On appelle *produit de deux événements* A et B et on note AB l'événement C consistant en la réalisation simultanée des événements A et B.

Par exemple, si l'événement A est « tirage d'un as » dans un jeu de cartes, l'événement B est « tirage d'un carreau », l'événement $C = AB$ correspond au «tirage d'un as de carreau ».

On appelle *produit de plusieurs événements* l'événement correspondant à la réalisation simultanée de tous ces événements.

Par exemple, si l'on tire trois coups sur une cible et que l'on considère les événements suivants:

B_1 — le premier coup est raté;
B_2 — le second coup est raté;
B_3 — le troisième coup est raté,

l'événement

$$B = B_1B_2B_3$$

correspond à ce qu'aucun coup n'a atteint le but.

Pour faciliter le calcul des probabilités il est d'usage de représenter des événements compliqués sous la forme de combinaisons d'événements plus simples en utilisant les opérations d'addition et de multiplication.

Par exemple, si trois coups sont tirés sur une cible et que l'on considère les événements élémentaires suivants:

A_1 — le premier coup atteint le but;
$\bar{A}_1$ — le premier coup est raté;
A_2 — le second coup atteint le but;
$\bar{A}_2$ — le second coup est raté;
A_3 — le troisième coup atteint le but;
$\bar{A}_3$ — le troisième coup est raté,

alors l'événement composé B correspondant à ce qu'exactement un coup atteint le but des trois coups tirés peut être présenté comme la combinaison des événements élémentaires:

$$B = A_1\bar{A}_2\bar{A}_3 + \bar{A}_1A_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3.$$

L'événement C selon lequel deux coups au moins atteindront la cible peut s'écrire comme suit:

$$C = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3 + A_1A_2A_3.$$

En théorie des probabilités on représente souvent ainsi les événements composés.

Par définition de la somme et du produit des événements on a

$$A + A = A;$$
$$AA = A.$$

Si l'événement B est un cas particulier de l'événement A on a:

$$A + B = A;$$
$$AB = B.$$

Les notions de *somme* et de *produit* des événements peuvent être illustrées par des images explicatives comme celles des figures 3.1.1 et 3.1.2. Désignons par A l'événement « le point tombe dans le

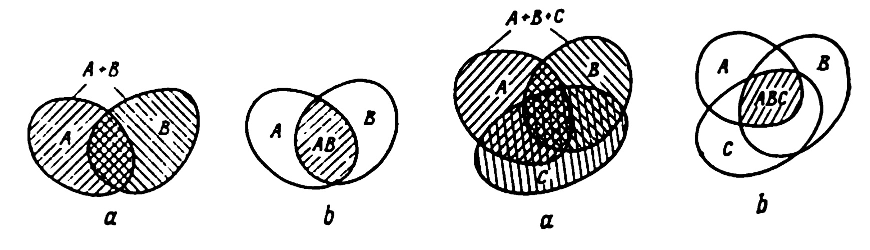


Fig. 3.1.1 Fig. 3.1.2

domaine A », et par B l'événement « le point tombe dans le domaine B », alors l'événement $A + B$ est que le point tombe dans le domaine hachuré sur la figure 3.1.1,*a* et l'événement AB est qu'il tombe dans le domaine hachuré sur la figure 3.1.1,*b*.

D'une manière analogue on a montré sur la figure 3.1.2 la somme et le produit de trois événements.

3.2. Théorème des probabilités totales

Ce théorème pose:

La probabilité de réaliser l'événement somme de deux événements incompatibles est égale à la somme des probabilités de ces événements:

$$P(A + B) = P(A) + P(B). \qquad (3.2.1)$$

Nous allons démontrer ce théorème pour le schéma du tirage exhaustif (système de « cas »). Supposons que les résultats éventuels d'une expérience sont n « cas », que nous figurons par n points:

$$\overbrace{\cdot\ \cdot\ \cdot\ \cdot}^{m \sim A}\ \overbrace{\cdot\ \cdot\ \cdot}^{k \sim B}\ \cdot\ \cdot\ \cdot$$
$$\underbrace{\qquad\qquad\qquad\qquad}_{n}$$

Supposons que parmi ces n cas m soient favorables à l'événement A, et k à l'événement B. On a alors

$$P(A) = \frac{m}{n}; \quad P(B) = \frac{k}{n}.$$

Les événements A et B étant incompatibles, aucun cas n'est favorable à la réalisation simultanée de A et B. Par conséquent, $m + k$ cas sont favorables à l'événement $A + B$ de sorte que

$$P(A + B) = \frac{m+k}{n}.$$

En substituant les expressions obtenues dans la formule (3.2.1) on obtient une identité. Le théorème se trouve ainsi démontré.

Généralisons ce théorème au cas de trois événements. Désignant par D l'événement $A + B$ et ajoutant à la somme encore un événement C, on montre facilement que

$$P(A + B + C) = P(D + C) = P(D) + P(C) =$$
$$= P(A + B) + P(C) = P(A) + P(B) + P(C).$$

Naturellement, ceci peut se généraliser par récurrence à un nombre quelconque d'événements incompatibles. En effet, supposons que ce théorème soit vrai pour n événements

$$A_1, A_2, \ldots, A_n;$$

montrons qu'il sera également vrai pour $n + 1$ événements

$$A_1, A_2, \ldots, A_n, A_{n+1}.$$

Notons:

$$A_1 + A_2 + \ldots + A_n = C.$$

Avec ces notations on a:

$$P(A_1 + A_2 + \ldots + A_n + A_{n+1}) = P(C + A_{n+1}) =$$
$$= P(C) + P(A_{n+1}).$$

Mais comme pour n événements le théorème est supposé démontré, on a:

$$P(C) = P(A_1) + P(A_2) + \ldots + P(A_n),$$

d'où

$$P(A_1 + A_2 + \ldots + A_n + A_{n+1}) =$$
$$= P(A_1) + P(A_2) + \ldots + P(A_n) + P(A_{n+1}),$$

ce qu'il fallait démontrer.

Ainsi le théorème des probabilités totales est applicable à un nombre quelconque d'événements incompatibles. Il est d'usage de

l'écrire comme suit:

$$P\left(\sum_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i). \tag{3.2.2}$$

Nous allons voir les corollaires qui en découlent.

Corollaire 1. *Si les événements $A_1, A_2, \ldots, A_n$ forment un système complet d'événements incompatibles, la somme de leurs probabilités est égale à l'unité:*

$$\sum_{i=1}^{n} P(A_i)=1.$$

Démonstration. Les événements $A_1, A_2, \ldots, A_n$ formant un système complet, la réalisation de l'un au moins d'entre eux est un événement certain:

$$P(A_1+A_2+\ldots+A_n)=1.$$

Les événements $A_1, A_2, \ldots, A_n$ étant incompatibles, on peut leur appliquer le théorème des probabilités totales:

$$P(A_1+A_2+\ldots+A_n)=P(A_1)+P(A_2)+\ldots+ +P(A_n)=\sum_{i=1}^{n} P(A_i),$$

d'où

$$\sum_{i=1}^{n} P(A_i)=1,$$

ce qu'il fallait démontrer.

Avant de démontrer le second corollaire du théorème des probabilités totales nous allons définir la notion d'événement contraire.

On appelle *événements contraires* deux événements incompatibles formant un système complet.

L'événement contraire de A est généralement désigné par $\bar{A}$.

Exemples d'événements contraires:

1) A — atteindre le but lors d'un coup;
   $\bar{A}$ — rater le coup;
2) B — obtenir face en jetant une pièce de monnaie;
   $\bar{B}$ — obtenir pile;
3) C — fonctionnement sans défaillance de tous les éléments d'un appareil;
   $\bar{C}$ — défaut au moins d'un élément;
4) D — trouver au moins deux pièces défectueuses dans un lot de contrôle;
   $\bar{D}$ — trouver au plus une pièce défectueuse.

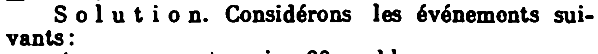
C o r o l l a i r e 2. *La somme des probabilités des événements contraires est égale à l'unité:*

$$P(A) + P(\bar{A}) = 1.$$

Ce corollaire est un cas particulier du corollaire 1. On l'a mis à part en raison de son importance pratique. Souvent dans la pratique il est plus facile de calculer la probabilité de l'événement contraire $\bar{A}$ que celle de l'événement A. Dans ces cas on calcule $P(\bar{A})$ et on trouve $P(A) = 1 - P(\bar{A})$.

Voyons quelques exemples d'application du théorème des probabilités totales et de ses corollaires.

E x e m p l e 1. Soient 1000 billets de loterie; un des billets gagne 500 roubles, 10 billets 100 roubles chacun. 50 billets gagnent 20 roubles chacun, 100 billets 5 roubles chacun, tous les autres ne gagnant rien. Trouver la probabilité pour le propriétaire d'un seul billet de gagner au moins 20 roubles.



S o l u t i o n. Considérons les événemonts suivants:

A — gagner au moins 20 roubles;
A_1 — gagner 20 roubles;
A_2 — gagner 100 roubles;
A_3 — gagner 500 roubles.

On a de toute évidence:

$$A = A_1 + A_2 + A_3.$$

En vertu du théorème des probabilités totales:

$$P(A) = P(A_1) + P(A_2) + P(A_3) = \\ = 0{,}050 + 0{,}010 + 0{,}001 = 0{,}061.$$

Fig. 3.2.1

E x e m p l e 2. Une cible ronde (fig. 3.2.1) a trois zones I, II et III. La probabilité d'atteindre la première zone avec un coup est 0,15; la seconde 0,23; la troisième 0,17. Trouver la probabilité de rater le coup.

S o l u t i o n. Désignons par A un coup raté et par $\bar{A}$ un coup atteignant le but. On a alors:

$$\bar{A} = \bar{A}_1 + \bar{A}_2 + \bar{A}_3,$$

où $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ sont les événements « atteindre la première, la seconde et la troisième zone », et

$$P(\bar{A}) = P(\bar{A}_1) + P(\bar{A}_2) + P(\bar{A}_3) = 0{,}15 + 0{,}23 + 0{,}17 = 0{,}55,$$

d'où

$$P(A) = 1 - P(\bar{A}) = 0{,}45.$$

E x e m p l e 3. Au bout d'un certain temps de fonctionnement un appareil peut se trouver dans l'un des états suivants:

1 — bon état de fonctionnement;
2 — légèrement déréglé;
3 — certaines pièces doivent être remplacées;
4 — une révision générale s'impose.

Les probabilités respectives de ces quatre états sont 0,55; 0,1; 0,15 et 0,2. Dans le premier cas l'appareil peut continuer à fonctionner, dans le second cas il faut 10 minutes de réglages, dans le troisième les réparations prennent

deux heures, dans le quatrième 24 heures. Trouver la probabilité pour qu'au bout de trois heures après la fin du service précédent l'appareil soit en mesure de fonctionner.

S o l u t i o n. Considérons les événements:

A — l'appareil est prêt à fonctionner après 3 heures;

A_1 — l'appareil peut fonctionner immédiatement;

A_2 — nécessité d'un réglage;

A_3 — nécessité de remplacer certaines pièces.

On a de toute évidence:

$$A = A_1 + A_2 + A_3.$$

En vertu du théorème des probabilités totales on a:

$$P(A) = P(A_1) + P(A_2) + P(A_3) = 0{,}55 + 0{,}1 + 0{,}15 = 0{,}8.$$

Pour résoudre ce problème il serait plus facile de passer à l'événement contraire $\bar{A}$ — «l'appareil ne sera pas prêt à fonctionner dans 3 heures».

L'événement $\bar{A}$ est équivalent à la nécessité d'une grosse réparation; d'où:

$$P(\bar{A}) = 0{,}2.$$
$$P(A) = 1 - P(\bar{A}) = 0{,}8.$$

Comme nous l'avons déjà noté, le théorème des probabilités totales (3.2.1) n'est vrai que pour des événements incompatibles.

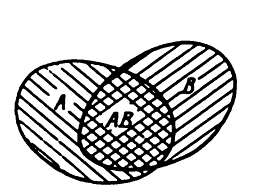


Fig. 3.2.2

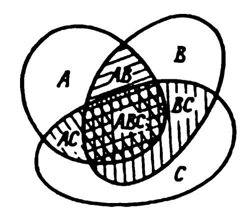


Fig. 3.2.3

Lorsque les événements A et B ne s'excluent pas mutuellement, la probabilité de l'événement somme est donnée par la formule:

$$P(A + B) = P(A) + P(B) - P(AB). \qquad (3.2.3)$$

La figure 3.2.2 donne une démonstration figurative de la formule 3.2.3.

D'une manière analogue, la probabilité de l'événement somme de trois événements simultanés est donnée par la formule:

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) -$$
$$- P(AC) - P(BC) + P(ABC).$$

La figure 3.2.3 permet de s'en convaincre.

On démontre aisément par récurrence la formule générale de la probabilité de la somme d'un nombre quelconque d'événements

simultanés:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i,j} (A_i A_j) +$$
$$+ \sum_{i,j,k} P(A_i A_j A_k) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n), \quad (3.2.4)$$

où les sommes sont étendues à différents indices i; i, j; i, j, k, etc.

La formule (3.2.4) exprime la probabilité de la somme d'un nombre quelconque d'événements en fonction des probabilités des produits de ces événements pris un à un, deux à deux, etc.

On peut écrire une formule analogue pour le produit des événements. En effet, la figure 3.2.2 montre que

$$P(AB) = P(A) + P(B) - P(A + B). \quad (3.2.5)$$

De même la figure 3.2.3 donne:

$$P(ABC) = P(A) + P(B) + P(C) - P(A + B) -$$
$$- P(A + C) - P(B + C) + P(A + B + C). \quad (3.2.6)$$

La formule générale exprimant la probabilité du produit d'un nombre quelconque d'événements en fonction des probabilités des sommes de ces événements pris un à un, deux à deux, etc., s'écrit:

$$P(A_1 A_2 \ldots A_n) = \sum_i P(A_i) - \sum_{i,j} P(A_i + A_j) +$$
$$+ \sum_{i,j,k} P(A_i + A_j + A_k) + \ldots + (-1)^{n-1} P(A_1 + \ldots + A_n). \quad (3.2.7)$$

Les formules du type (3.2.4) et (3.2.7) sont couramment utilisées pour les transformations des différentes expressions contenant les probabilités des sommes et des produits d'événements. Dans certains problèmes il est plus commode d'utiliser seulement des sommes et dans d'autres seulement des produits; des formules de transformation permettent de passer des unes aux autres.

Exemple. Un appareil se compose de trois dispositifs, de deux dispositifs du premier type: A_1 et A_2, et d'un dispositif du second type B. Les dispositifs A_1 et A_2 sont interchangeables, lorsque l'un est en défaut le second se met automatiquement en fonctionnement. Le dispositif B fonctionne indépendamment. Pour que l'appareil cesse de fonctionner il faut que les deux dispositifs A_1 et A_2 ou le dispositif B soient défectueux. Ainsi le défaut de l'appareil, l'événement C, peut s'écrire comme suit:

$$C = A_1 A_2 + B,$$

où A_1 est le défaut du dispositif A_1; A_2 le défaut du dispositif A_2; B le défaut du dispositif B.

Exprimer la probabilité de l'événement C en fonction des probabilités des événements qui soient des sommes et non des produits des événements élémentaires A_1, A_2 et B.

Solution. Utilisons la formule (3.2.3):

$$P(C) = P(A_1A_2) + P(B) - P(A_1A_2B). \qquad (3.2.8)$$

puis la formule (3.2.5):

$$P(A_1A_2) = P(A_1) + P(A_2) - P(A_1 + A_2)$$

et enfin la formule (3.2.6):

$$P(A_1A_2B) = P(A_1) + P(A_2) + P(B) - P(A_1 + A_2) - \\ - P(A_1 + B) - P(A_2 + B) + P(A_1 + A_2 + B).$$

En substituant ces expressions dans (3.2.8) et en simplifiant on obtient:

$$P(C) = P(A_1 + B) + P(A_2 + B) - P(A_1 + A_2 + B).$$

3.3. Théorème des probabilités composées

Avant d'exposer ce théorème introduisons la notion importante d'événements indépendants et dépendants.

L'événement A est dit *indépendant de l'événement* B si la probabilité de A ne dépend pas de ce que B se produit ou non.

L'événement A est *dépendant de l'événement* B si la probabilité de A change suivant que B s'est produit ou non.

Soient quelques exemples.

1) Une expérience consiste à jeter deux pièces de monnaie; considérons les événements:

A — obtenir pile sur la première pièce:
B — obtenir pile sur la seconde pièce.

Dans ce cas la probabilité de A ne dépend pas de la réalisation de B, donc l'événement A est indépendant de l'événement B.

2) Une urne contient deux boules blanches et une noire; deux personnes tirent chacune une boule de l'urne; considérons les événements:

A — la première personne a tiré une boule blanche;
B — la seconde personne a tiré une boule blanche.

La probabilité de l'événement A en l'absence de l'information sur l'événement B est égale à 2/3. Si l'on sait que l'événement B s'est réalisé la probabilité de A devient égale à 1/2, d'où on conclut que A dépend de B.

On appelle *probabilité conditionnelle de* A *relative à* B et on note $P(A \mid B)$ la probabilité de l'événement A calculée sous la condition que l'événement B s'est réalisé. Pour les conditions du dernier exemple on a:

$$P(A) = \frac{2}{3}; \qquad P(A \mid B) = \frac{1}{2}.$$

On peut indiquer que A est indépendant de B en écrivant:

$$P(A \mid B) = P(A),$$

la condition de dépendance s'exprime:

$$P(A \mid B) \neq P(A).$$

Nous pouvons maintenant formuler et démontrer le théorème des probabilités composées.

La probabilité du produit de deux événements est égale au produit de la probabilité de l'un d'eux par la probabilité conditionnelle de l'autre, calculée sous la condition que le premier ait lieu:

$$P(AB) = P(A)\,P(B \mid A). \qquad (3.3.1)$$

Nous allons démontrer ce théorème pour des événements formant un système complet (système de « cas »). Supposons que les résultats éventuels de l'expérience sont n « cas », que nous figurons comme auparavant par n points:

$$\underbrace{\overbrace{\underbrace{\overbrace{\cdots\cdots}^{m\sim A}\cdots}_{l\sim AB}}^{k\sim B}\cdots}_{n}$$

Supposons que m cas soient favorables à l'événement A et k cas à l'événement B. Comme nous n'avons pas supposé les événements A et B incompatibles nous devons admettre qu'il existe en général des cas favorables à l'événement A et à l'événement B simultanément. Soit l le nombre de tels cas. On a alors:

$$P(AB) = \frac{l}{n}; \qquad P(A) = \frac{m}{n}.$$

Calculons $P(B \mid A)$, c'est-à-dire la probabilité conditionnelle de l'événement B en supposant que A soit réalisé. Si l'on sait que l'événement A a eu lieu, seuls m cas, notamment les cas favorables à l'événement A, sont à considérer au lieu des n dont on disposait au début de l'expérience, l de ces m cas étant favorables à l'événement B. Par conséquent:

$$P(B \mid A) = \frac{l}{m}.$$

En substituant les expressions de $P(AB)$, $P(A)$ et $P(B \mid A)$ dans la formule (3.3.1) on obtient une identite, ce qui démontre le théorème.

Il est clair que le théorème des probabilités composées est indifférent à l'ordre des événements A et B, ce qui permet d'écrire ce théorème comme suit:

$$P(AB) = P(B)\,P(A \mid B).$$

Mentionnons les corollaires du théorème des probabilités composées.

Corollaire 1. *Si A est indépendant de B, alors B ne dépend pas de A.*

Démonstration. On sait que l'événement A ne dépend pas de B, c'est-à-dire que

$$P(A) = P(A \mid B). \tag{3.3.2}$$

Il y a lieu de démontrer que B ne dépend pas de A, c'est-à-dire que

$$P(B) = P(B \mid A).$$

Pour la démonstration nous supposons que $P(A) \neq 0$.

Ecrivons le théorème des probabilités composées sous ses deux formes:

$$P(AB) = P(A)\,P(B \mid A);$$
$$P(AB) = P(B)\,P(A \mid B),$$

d'où

$$P(A)\,P(B \mid A) = P(B)\,P(A \mid B)$$

ou en vertu de la condition (3.3.2)

$$P(A)\,P(B \mid A) = P(B)\,P(A). \tag{3.3.3}$$

Divisons les deux membres de (3.3.3) par $P(A)$. On a:

$$P(B \mid A) = P(B),$$

ce qu'il fallait démontrer.

En vertu du corollaire 1 la dépendance ou l'indépendance des événements sont toujours *réciproques*. Ceci permet de donner une nouvelle définition des événements indépendants.

Deux événements sont dits indépendants si la réalisation de l'un d'eux ne change pas la probabilité de réalisation de l'autre.

La notion d'indépendance des événements peut être étendue à un nombre quelconque d'événements. Plusieurs événements sont dits mutuellement indépendants si aucun d'eux ne dépend de toute combinaison des autres.

Corollaire 2. *La probabilité du produit de deux événements indépendants est égale au produit des probabilités de ces événements.*

Le corollaire découle immédiatement de la définition des événements indépendants.

Le théorème des probabilités composées peut être généralisé à un nombre quelconque d'événements. Sous sa forme générale, il peut se formuler comme suit:

La probabilité du produit de plusieurs événements est égale au produit des probabilités de ces événements, la probabilité de chaque i-ème événement étant calculée sous la condition que tous les $i - 1$

événements se soient produits:

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots$$
$$\ldots P(A_n|A_1A_2\ldots A_{n-1}). \qquad (3.3.4)$$

Cette formule peut être démontrée par récurrence.

Pour les événements indépendants le théorème se simplifie et s'écrit comme suit:

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2)\ldots P(A_n), \qquad (3.3.5)$$

c'est-à-dire que la *probabilité du produit des événements indépendants est égale au produit des probabilités de ces événements.*

En utilisant le signe du produit on peut écrire ce théorème comme suit:

$$P\left(\prod_{i=1}^{n} A_i\right)=\prod_{i=1}^{n} P(A_i). \qquad (3.3.6)$$

Voyons quelques exemples d'application du théorème des probabilités composées.

Exemple 1. Une urne contient 2 boules blanches et 3 noires. On tire deux boules de l'urne. Trouver la probabilité pour que les deux boules tirées soient blanches.

Solution. Nous allons désigner par A l'apparition de deux boules blanches. L'événement A est le produit de deux événements élémentaires: A_1 — « la première boule tirée est blanche »; A_2 — « la seconde boule tirée est blanche »:

$$A=A_1A_2.$$

En vertu du théorème des probabilités composées on a:

$$P(A)=P(A_1)P(A_2|A_1)=\frac{2}{5}\cdot\frac{1}{4}=0{,}1.$$

Exemple 2. Même problème que précédemment, mais avec la remise de la boule après le premier tirage.

Solution. Dans ce cas les événements A_1 et A_2 sont indépendants et on a:

$$P(A)=P(A_1)P(A_2)=\frac{2}{5}\cdot\frac{2}{5}=0{,}16.$$

Exemple 3. Un appareil se compose de trois dispositifs. Durant le temps t de fonctionnement de l'appareil les défauts des trois dispositifs peuvent survenir de façon indépendante. Un seul défaut suffit pour mettre l'appareil hors service. Pour ce temps t la fiabilité (probabilité de fonctionnement sans défaillance) du premier dispositif est égale à $p_1=0{,}8$; du second à $p_2=0{,}9$; du troisième à $p_3=0{,}7$. Trouver la fiabilité de l'appareil.

Solution. Notons:

A — fonctionnement sans défaillance de l'appareil;
A_1 — fonctionnement sans défaillance du premier dispositif;
A_2 — fonctionnement sans défaillance du second dispositif;
A_3 — fonctionnement sans défaillance du troisième dispositif;

avec ces notations on a:

$$A=A_1A_2A_3,$$

d'où en vertu du théorème des probabilités composées pour des événements indépendants:

$$P(A) = P(A_1)\,P(A_2)\,P(A_3) = p_1p_2p_3 = 0{,}504.$$

Très rarement dans les applications pratiques on peut se borner à l'utilisation exclusive de l'un ou l'autre des théorèmes examinés. Généralement, ces deux théorèmes sont utilisés ensemble, l'événement dont il y a lieu de trouver la probabilité étant représenté comme une somme de plusieurs événements incompatibles (variantes de l'événement étudié), tandis que chacun d'eux est à son tour un produit d'événements.

E x e m p l e 4. Trois coups sont tirés sur une même cible. Les probabilités d'atteindre la cible lors du premier, du second et du troisième coup sont respectivement:

$$p_1 = 0.4;\quad p_2 = 0.5;\quad p_3 = 0.7.$$

Trouver la probabilité pour que de ces trois coups exactement l'un atteigne la cible.

S o l u t i o n. Considérons l'événement A — « exactement un coup atteint la cible ». Cet événement peut se produire de plusieurs manières différentes, c'est-à-dire qu'il se décompose en plusieurs variantes incompatibles.

Désignant A_1, A_2, A_3 les événements « atteindre la cible » lors du premier, du second et du troisième coup respectivement et

$\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ les événements « rater le but » lors du premier, le second et le troisième coup, on peut représenter ces variantes comme suit:

$$A_1\bar{A}_2\bar{A}_3;\ \bar{A}_1A_2\bar{A}_3;\ \bar{A}_1\bar{A}_2A_3,\ \text{et}\ A = A_1\bar{A}_2\bar{A}_3 + \bar{A}_1A_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3.$$

En utilisant les théorèmes des probabilités totales et composées et compte tenu de la propriété des événements contraires on trouve:

$$P(A) = P(A_1\bar{A}_2\bar{A}_3) + P(\bar{A}_1A_2\bar{A}_3) + P(\bar{A}_1\bar{A}_2A_3) =$$
$$= 0{,}4\cdot 0{,}5\cdot 0{,}3 + 0{,}6\cdot 0{,}5\cdot 0{,}3 + 0{,}6\cdot 0{,}5\cdot 0{,}7 = 0{,}36.$$

E x e m p l e 5. Dans les conditions de l'exemple précédent trouver la probabilité pour qu'au moins un coup atteigne la cible.

S o l u t i o n. Soit l'événement B — « au moins un coup atteint la cible ». En utilisant les notations et la méthode de l'exemple précédent on peut écrire l'événement B comme une somme de variantes incompatibles, soit:

$$B = A_1A_2A_3 + A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3 + A_1\bar{A}_2\bar{A}_3 +$$
$$+ \bar{A}_1A_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3,$$

et à l'aide du théorème des probabilités composées trouver la probabilité de chacune des variantes et les additionner. Mais cette voie est trop longue; il est bon ici de passer de l'événement B à son contraire $\bar{B}$: « aucun coup n'a atteint la cible ». On a de toute évidence:

$$\bar{B} = \bar{A}_1\bar{A}_2\bar{A}_3.$$

En vertu du théorème des probabilités composées

$$P(\bar{B}) = P(\bar{A}_1\bar{A}_2\bar{A}_3) = 0{,}6\cdot 0{,}5\cdot 0{,}3 = 0{,}09,$$

d'où

$$P(B) = 1 - P(\bar{B}) = 1 - 0{,}09 = 0{,}91.$$

Le dernier exemple est une illustration de l'utilité des événements contraires en théorie des probabilités. On peut se guider du principe suivant : *si l'événement contraire $\bar{A}$ se décompose en un nombre moindre de variantes que l'événement A, il y a lieu pour calculer les probabilités de passer à l'événement contraire.*

E x e m p l e 6. Une urne contient 10 boules blanches et 5 noires; on tire 4 boules de l'urne. Trouver la probabilité pour que l'une des boules tirées soit noire.

S o l u t i o n. L'événement A : « tirer au moins une boule noire » a un grand nombre de variantes (parmi les boules tirées il peut y avoir des combinaisons différentes de boules noires et blanches). Il est bien plus simple de passer à l'événement contraire $\bar{A}$: « parmi les boules tirées aucune n'est noire ».

L'événement $\bar{A}$ est le produit de quatre événements:

$$\bar{A} = B_1B_2B_3B_4$$

dont les significations sont les suivantes:

B_1 — la première boule est blanche;
B_2 — la seconde boule est blanche;
B_3 — la troisième boule est blanche;
B_4 — la quatrième boule est blanche.

En vertu du théorème des probabilités composées on a:

$$P(\bar{A}) = P(B_1)\,P(B_2 \mid B_1)\cdot P(B_3 \mid B_1B_2) \times$$
$$\times P(B_4 \mid B_1B_2B_3) = \frac{10}{15}\cdot\frac{9}{14}\cdot\frac{8}{13}\cdot\frac{7}{12} = \frac{2}{13},$$

d'où

$$P(A) = 1 - P(\bar{A}) = \frac{11}{13}.$$

E x e m p l e 7. Un appareil se compose de trois éléments. La fiabilité (probabilité de fonctionnement sans défaillance durant un intervalle de temps donné) du premier élément est égale à 0,9 ; du second à 0,8 ; du troisième à 0,7. Le premier élément est indispensable au fonctionnement de l'appareil; si le second ou le troisième élément est en défaut, l'appareil fonctionne mais avec un rendement inférieur; la défaillance simultanée du second et du troisième élément fait que l'appareil ne peut fonctionner. Les éléments deviennent défectueux indépendamment les uns des autres. Trouver la probabilité pour que l'appareil fonctionne durant un intervalle de temps donné (même avec un rendement inférieur).

S o l u t i o n. Désignons par

A_1 — le fonctionnement sans défaillance du premier élément;
A_2 — le fonctionnement sans défaillance du second élément;
A_3 — le fonctionnement sans défaillance du troisième élément.

Les événements $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ sont respectivement les défaillances du premier, du second et du troisième élément. L'événement A (l'appareil reste en service, même avec un rendement inférieur) peut s'écrire comme suit:

$$A = A_1A_2A_3 + A_1\bar{A}_2A_3 + A_1A_2\bar{A}_3. \tag{3.3.7}$$

L'événement contraire $\bar{A}$ (l'appareil ne fonctionne pas) s'écrit:

$$\bar{A} = \bar{A}_1A_2A_3 + \bar{A}_1\bar{A}_2A_3 + \bar{A}_1A_2\bar{A}_3 + A_1\bar{A}_2\bar{A}_3 + \bar{A}_1\bar{A}_2\bar{A}_3.$$

Ici l'événement contraire $\overline{A}$ a plus de variantes que l'événement A, donc il n'y a aucun avantage de l'utiliser. En développant la formule (3.3.7) en conformité des théorèmes des probabilités totales et composées on a:

$$P(A) = P(A_1A_2A_3) + P(A_1\overline{A}_2A_3) + P(A_1A_2\overline{A}_3) =$$

$$= P(A_1)P(A_2)P(A_3) + P(A_1)P(\overline{A}_2)P(A_3) + P(A_1)P(A_2)P(\overline{A}_3) =$$

$$= 0{,}9\cdot 0{,}8\cdot 0{,}7 + 0{,}9\cdot 0{,}2\cdot 0{,}7 + 0{,}9\cdot 0{,}8\cdot 0{,}3 = 0{,}846.$$

Exemple 8. Un train se compose de 10 wagons; chacun des wagons peut avec une probabilité 0,1 (indépendamment des autres) avoir un défaut. Avant le départ les wagons sont examinés indépendamment les uns des autres par deux contrôleurs, chacun décelant le défaut (s'il y a eu un) avec une probabilité égale à 0,7. Un seul défaut décelé suffit pour que le train so tretardé. Trouver les probabilités des événements suivants:

A — le train sera retardé;
B — le train partira avec au moins un wagon défectueux.

Solution. L'événement A ayant un grand nombre de variantes, et il est plus commode de passer à l'événement contraire $\overline{A}$: le train ne sera pas retardé (aucun défaut n'est décelé).

L'événement $\overline{A}$ est le produit de 10 événements que nous désignons par C consistant en ce qu'un wagon donné n'aura pas de défaut décelé. L'événement C se réalise par deux manières:

$$C = C_1 + C_2,$$

où

C_1 — un wagon quelconque est sans défaut;
C_2 — le wagon a un défaut non decelé, ni par le premier, ni par le second contrôleur.

On a:

$$P(C) = P(C_1) + P(C_2),$$
$$P(C_1) = 0{,}9,$$
$$P(C_2) = 0{,}1\cdot 0{,}3\cdot 0{,}3 = 0{,}009,$$
$$P(C) = 0{,}909 \text{ et}$$
$$P(\overline{A}) = 0{,}909^{10} \approx 0{,}385; \quad P(A) = 1 - P(\overline{A}) \approx 0{,}615.$$

Pour calculer la probabilité de l'événement B il y a lieu de noter que l'événement $\overline{A}$ se décompose en deux variantes:

$$\overline{A} = D + B,$$

où D — le train n'a pas de wagon défectueux.

D'où:

$$P(B) = P(\overline{A}) - P(D).$$

L'événement D est le produit de 10 événements consistant en ce qu'un wagon quelconque n'a pas de défaut, alors:

$$P(D) = 0{,}9^{10} \approx 0{,}349,$$
$$P(B) = 0{,}385 - 0{,}349 = 0{,}036.$$

Exemple 9. On jette une pièce de monnaie 6 fois. Trouver la probabilité pour que face sorte plus fréquemment que pile.

Solution. Pour trouver la probabilité de l'événement qui nous intéresse (nombre de faces supérieur à celui de piles) on pourrait énumérer toutes

ses variantes possibles, par exemple:

A_1 — six faces et aucun pile,

A_2 — cinq faces et un pile,

etc.

Mais il est plus simple d'utiliser une autre méthode. Enumérons toutes les issues éventuelles de l'expérience:

A — face sort plus fréquemment que pile;

B — pile sort plus fréquemment que face;

C — pile et face sortent en nombre égal.

Les événements A, B, C sont incompatibles et forment un système complet d'événements. Par conséquent:

$$P(A) + P(B) + P(C) = 1.$$

Le problème étant symétrique par rapport à pile et face, on a:

$$P(A) = P(B),$$

d'où

$$2P(A) + P(C) = 1$$

et

$$P(A) = \frac{1-P(C)}{2}.$$

Cherchons la probabilité de l'événement C de réaliser en six jets exactement trois fois face (et donc trois fois pile). La probabilité de chacune des variantes de l'événement C (par exemple, de la suite face, pile, face, face, pile, pile) est la même et égale à $(1/2)^6$. Le nombre de ces combinaisons étant égal à $C_6^3 = 20$ (nombre de manières permettant de choisir des six jets les trois où face apparaît), on a:

$$P(C) = \frac{20}{64} = \frac{5}{16},$$

d'où

$$P(A) = \frac{1}{2}\left(1 - \frac{5}{16}\right) = \frac{11}{32}.$$

E x e m p l e 10. Un appareil se compose de quatre éléments: A_1, A_2, A_3, A_4, l'élément A_2 doublant A_1, et A_4 doublant A_3. Lorsque l'un des éléments principaux (A_1 ou A_3) est en panne, on passe automatiquement à l'élément auxiliaire. La fiabilité (probabilité de fonctionnement sans défaillance) de chacun des éléments pour un certain laps de temps est respectivement p_1, p_2, p_3, p_4. La fiabilité des dispositifs de commutation est égale à p. Les défauts des éléments surviennent indépendamment. Trouver la fiabilité de l'appareil.

S o l u t i o n. Considérons deux ensembles généralisés B et C englobant l'un les éléments A_1, A_2 et le dispositif de commutation, et l'autre les éléments A_3, A_4 et le dispositif de commutation correspondant. Soient les événements:

A — fonctionnement sans défaillance de l'appareil,

B — fonctionnement sans défaillance de l'ensemble B,

C — fonctionnement sans défaillance de l'ensemble C.

Il est clair que

$$A = BC,$$

d'où

$$P(A) = P(B) \cdot P(C).$$

Trouvons la probabilité de l'événement B. Celui-ci se décompose en deux variantes:

A_1 — l'élément A_1 fonctionne sans défaillance,
A_2' — l'élément A_1 se trouve défectueux mais le dispositif de commutation et l'élément A_2 fonctionnent normalement.

On a:

$$P(B) = P(A_1) + P(A_2') = p_1 + (1 - p_1)\, pp_2,$$

d'une manière analogue:

$$P(C) = p_3 + (1 - p_3)\, pp_4.$$

d'où:

$$P(A) = [p_1 + (1 - p_1)\, pp_2]\,[p_3 + (1 - p_3)\, pp_4].$$

3.4. Formule des probabilités totales

Cette formule est un corollaire des deux théorèmes fondamentaux: du théorème des probabilités totales et du théorème des probabilités composées.

Soit à déterminer la probabilité d'un certain événement A pouvant avoir lieu simultanément avec l'un des événements

$$H_1, H_2, \ldots, H_n,$$

formant un système complet d'événements incompatibles. Nous appellerons ces événements *hypothèses.*

Montrons que dans le cas général on a:

$$P(A) = \sum_{i=1}^{n} P(H_i)\, P(A \mid H_i), \tag{3.4.1}$$

c'est-à-dire que la probabilité de l'événement A se calcule comme la somme des produits de la probabilité de chacune des hypothèses par la probabilité de l'événement conditionnellement à cette hypothèse.

La formule (3.4.1) est appelée *formule des probabilités totales.*

D é m o n s t r a t i o n. Les hypothèses $H_1, H_2, \ldots, H_n$ formant un système complet, l'événement A ne se produit qu'en combinaison avec l'une de ces hypothèses:

$$A = H_1A + H_2A + \ldots + H_nA.$$

Les hypothèses $H_1, H_2, \ldots, H_n$ étant incompatibles, les combinaisons $H_1A, H_2A, \ldots, H_nA$ le sont également; en leur appliquant le théorème des probabilités totales on obtient:

$$P(A) = P(H_1A) + P(H_2A) + \ldots + P(H_nA) = \sum_{i=1}^{n} P(H_i\ ').$$

En appliquant aux événements H_iA le théorème des probabilités composées on obtient:

$$P(A)=\sum_{i=1}^{n} P(H_i)P(A|H_i),$$

ce qu'il fallait démontrer.

Exemple 1. Soient trois urnes identiques, la première contenant deux boules blanches et une noire; la seconde trois boules blanches et une noire et la troisième deux boules blanches et deux noires. On choisit une urne au hasard et on en tire une boule. Trouver la probabilité de tirer une boule blanche.

Solution. Considérons les trois hypothèses suivantes:

H_1 — on choisit la première urne;
H_2 — on choisit la seconde urne;
H_3 — on choisit la troisième urne

et l'événement A : « tirage d'une boule blanche ».

Les hypothèses étant équiprobables, on a:

$$P(H_1) = P(H_2) = P(H_3) = 1/3.$$

Les probabilités conditionnelles de l'événement A sous ces hypothèses sont respectivement:

$$P(A \mid H_1) = 2/3; \quad P(A \mid H_2) = 3/4; \quad P(A \mid H_3) = 1/2.$$

La formule des probabilités totales conditionnelles donne:

$$P(A)=\frac{1}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{3}{4}+\frac{1}{3}\cdot\frac{1}{2}=\frac{23}{36}.$$

Exemple 2. Douze appareils sont en exploitation: trois d'entre eux sont fabriqués par l'usine n° 1, quatre par l'usine n° 2 et cinq par l'usine n° 3. Les appareils provenant de l'usine n° 1 passent l'essai avec une probabilité 0,9, de l'usine n° 2 avec une probabilité 0,8, de l'usine n° 3 avec une probabilité 0,75. Trouver la probabilité pour qu'un appareil choisi au hasard passe l'essai.

Solution. Nous allons envisager les hypothèses suivantes:

H_1 — l'appareil choisi provient de l'usine n° 1;
H_2 — l'appareil choisi provient de l'usine n° 2;
H_3 — l'appareil choisi provient de l'usine n° 3.

Les probabilités de ces hypothèses sont:

$$P(H_1) = 3/12 = 1/4,$$
$$P(H_2) = 4/12 = 1/3,$$
$$P(H_3) = 5/12.$$

Sous ces hypothèses, les probabilités conditionnelles de l'événement A : « l'appareil a passé l'essai » sont:

$$P(A \mid H_1) = 0,9;$$
$$P(A \mid H_2) = 0,8;$$
$$P(A \mid H_3) = 0,75.$$

En utilisant la formule des probabilités totales conditionnelles on obtient:

$$P(A) = 1/4\cdot 0,9 + 1/3\cdot 0,8 + 5/18\cdot 0,75 = 0,804.$$

Exemple 3. Deux régulateurs contrôlent le fonctionnement d'un moteur. Il est désirable que durant un temps t le moteur fonctionne sans panne. En présence des deux régulateurs la panne du moteur peut survenir avec une probabilité égale à $q_{1,2}$; lorsque seul le premier régulateur fonctionne, avec une probabilité égale à q_1, et lorsque les deux régulateurs sont en panne, avec une probabilité égale à q_2. La fiabilité du premier régulateur est égale à P_1, celle du second à P_2. Les éléments se mettant en panne indépendamment les uns des autres, trouver la fiabilité totale (probabilité de fonctionnement sans panne) du moteur.

Solution. Considérons les hypothèses suivantes:

$H_{1,2}$ — les deux régulateurs fonctionnent;

H_1 — seul le premier régulateur fonctionne (le second est en panne);

H_2 — seul le second régulateur fonctionne (le premier est en panne);

H_0 — les deux régulateurs sont en panne.

et l'événement A: « l'appareil fonctionne sans panne ».

Les probabilités des hypothèses énumérées sont:

$$P(H_{1,2}) = P_1P_2; \quad P(H_1) = P_1(1-P_2);$$

$$P(H_2) = P_2(1-P_1); \quad P(H_0) = (1-P_1)(1-P_2).$$

Les probabilités conditionnelles de l'événement A sous ces conditions sont:

$$P(A \mid H_{1,2}) = 1-q_{1,2}; \quad P(A \mid H_1) = 1-q_1;$$

$$P(A \mid H_2) = 1-q_2; \quad P(A \mid H_0) = 1-q_0.$$

Selon la formule des probabilités totales conditionnelles:

$$P(A) = P_1P_2(1-q_{1,2}) + P_1(1-P_2)(1-q_1) + \\ + P_2(1-P_1)(1-q_2)(1-P_1)(1-P_2)(1-q_0).$$

3.5. Théorème des hypothèses (formule de Bayes)

Le *théorème des hypothèses* ou *formule de Bayes* est un corollaire du théorème des probabilités totales et de la formule des probabilités totales conditionnelles.

Nous allons envisager le problème suivant.

Soit un système complet d'hypothèses incompatibles $H_1, H_2, \ldots, H_n$. Les probabilités de ces hypothèses avant l'expérience sont données et sont respectivement $P(H_1), P(H_2), \ldots, P(H_n)$. L'expérience a réalisé un certain événement A. Demandons-nous comment la réalisation de cet événement change les probabilités de ces hypothèses.

En fait il faut ici trouver la probabilité conditionnelle $P(H_i \mid A)$ pour chacune des hypothèses.

En vertu du théorème des probabilités composées on a:

$$P(AH_i) = P(A)P(H_i \mid A) = P(H_i)P(A \mid H_i) \quad (i=1, 2, \ldots, n),$$

ou, en négligeant le premier membre:

$$P(A)P(H_i \mid A) = P(H_i)P(A \mid H_i) \quad (i = 1, 2, \ldots, n),$$

d'où

$$P(H_i \mid A) = \frac{P(H_i)\,P(A \mid H_i)}{P(A)} \qquad (i = 1, 2, \ldots, n).$$

En exprimant $P(A)$ selon la formule (3.4.1) on obtient:

$$P(H_i \mid A) = \frac{P(H_i)\,P(A \mid H_i)}{\sum_{i=1}^{n} P(H_i)P(A \mid H_i)} \qquad (i = 1, 2, \ldots, n). \qquad (3.5.1)$$

La formule (3.5.1) est appelée *formule de Bayes* ou *théorème des hypothèses.*

Exemple 1. Un appareil peut être monté avec des pièces de haute qualité et avec des pièces ordinaires. Dans le premier cas sa fiabilité (probabilité de fonctionnement sans défaillance) durant le temps t est égale à 0,95; dans le second, à 0,7. Environ 40 % des appareils sont fabriqués avec des pièces de haute qualité. L'appareil a été soumis à l'essai pendant le temps t et s'est avéré bon. Trouver la probabilité qu'il soit monté avec des pièces de haute qualité.

Solution. Deux hypothèses suivantes sont possibles:

H_1 — l'appareil est monté avec des pièces de haute qualité;

H_2 — l'appareil est monté avec des pièces de qualité ordinaire.

Les probabilités de ces hypothèses avant l'expérience sont:

$$P(H_1) = 0.4; \quad P(H_2) = 0.6.$$

L'expérience a donné lieu à l'apparition de l'événement A : « fonctionnement sans défaillance durant le temps t ».

Les probabilités conditionnelles de cet événement sous les hypothèses H_1 et H_2 sont:

$$P(A \mid H_1) = 0{,}95; \quad P(A \mid H_2) = 0{,}7.$$

La formule (3.5.1) permet de trouver la probabilité de l'hypothèse H_1 après l'expérience:

$$P(H_1) = \frac{0{,}4 \cdot 0{,}95}{0.4 \cdot 0{,}95 + 0{,}6 \cdot 0{,}6} = 0{,}475.$$

Exemple 2. Deux tireurs tirent indépendamment l'un de l'autre sur une cible un coup chacun. La probabilité d'atteindre la cible est égale à 0,8 pour le premier tireur et 0.4 pour le second. Un coup a percé la cible. Trouver la probabilité pour que ce soit le coup du premier tireur.

Solution. Avant l'expérience les hypothèses suivantes sont possibles:

H_1 — aucun des deux tireurs n'atteint la cible;

H_2 — les deux tireurs atteignent la cible;

H_3 — seul le premier tireur atteint la cible;

H_4 — seul le second tireur atteint la cible.

Les probabilités de ces hypothèses sont:

$$P(H_1) = 0{,}2 \cdot 0{,}6 = 0{,}12;$$

$$P(H_2) = 0{,}8 \cdot 0{,}4 = 0{,}32;$$

$$P(H_3) = 0{,}8 \cdot 0{,}6 = 0{,}48;$$

$$P(H_4) = 0{,}2 \cdot 0{,}4 = 0{,}08.$$

Les probabilités conditionnelles de l'événement observé A sous ces hypothèses sont:

$$P(A \mid H_1) = 0; \quad P(A \mid H_2) = 0; \quad P(A \mid H_3) = 1; \quad P(A \mid H_4) = 1.$$

Après l'expérience les hypothèses H_1 et H_2 deviennent impossibles, et les probabilités des hypothèses H_3 et H_4 sont:

$$P(H_3 \mid A) = \frac{0,48 \cdot 1}{0,48 \cdot 1 + 0,08 \cdot 1} = \frac{6}{7}; \quad P(H_4 \mid A) = \frac{0,08 \cdot 1}{0,48 \cdot 1 + 0,08 \cdot 1} = \frac{1}{7}.$$

Par conséquent, avec une probabilité égale à $\frac{6}{7}$ on peut attribuer le trou au premier tireur.

Exemple 3. Deux stations d'observation fournissent des données sur un certain système qui peut se trouver dans deux états S_1 et S_2, en passant d'une manière aléatoire de l'un à l'autre. De longues observations ont permis d'établir que durant environ 30 % du temps le système se trouve dans l'état S_1, et 70 % dans l'état S_2. La station nº 1 fournit des données erronées environ dans 2 % de tous les cas, et la station nº 2 dans 8 %. A un certain moment la station nº 1 a communiqué que le système se trouve dans l'état S_1, et la station nº 2 qu'il est dans l'état S_2.

Laquelle des communications doit être supposée exacte?

Solution. Il est naturel de supposer vraie celle des communications dont la probabilité d'erreur est moindre. Nous allons utiliser la formule de Bayes. A cet effet faisons les hypothèses suivantes sur l'état du système:

H_1 — le système se trouve dans l'état S_1;
H_2 — le système se trouve dans l'état S_2.

L'événement observé A correspond à ce que la station nº 1 a transmis que le système se trouve dans l'état S_1, et la station nº 2 qu'il se trouve dans l'état S_2.

Les probabilités des hypothèses avant l'expérience sont:

$$P(H_1) = 0,3; \quad P(H_2) = 0,7.$$

Calculons les probabilités conditionnelles de l'événement observé A pour ces hypothèses. Pour que l'événement A ait lieu sous l'hypothèse H_1 il faut que la communication transmise par la première station soit vraie et par la seconde erronée:

$$P(A \mid H_1) = 0,98 \cdot 0,08 = 0,0784.$$

D'une manière analogue on a:

$$P(A \mid H_2) = 0,92 \cdot 0,02 = 0,0184.$$

En appliquant la formule de Bayes on peut trouver la probabilité pour que l'état réel du système soit S_1:

$$P(H_1 \mid A) = \frac{0,3 \cdot 0,0784}{0,3 \cdot 0,0784 + 0,7 \cdot 0,0184} \approx 0,645,$$

c'est-à-dire que des deux communications celle de la première station est plus vraisemblable.

CHAPITRE 4

EXPÉRIENCES RÉPÉTÉES

4.1. Théorème particulier sur les expériences répétées

Dans les applications pratiques de la théorie des probabilités on rencontre souvent des cas où une même expérience ou des expériences analogues sont répétées de nombreuses fois. Chaque expérience peut réaliser ou ne pas réaliser un événement A. Ce qui nous intéresse ce n'est pas le résultat de chaque expérience, mais le nombre de réalisations de l'événement A dans une série d'expériences. Par exemple, dans une série de coups tirés sur une cible, ce n'est pas le résultat de chaque coup qui nous intéresse en général, mais le nombre total d'atteintes. Dans des problèmes de ce genre il y a lieu de savoir trouver la probabilité d'un nombre quelconque de réalisations d'un événement dans une série d'expériences. Le présent chapitre est consacré à l'étude de ces problèmes. La solution en est assez simple dans le cas où les expériences sont indépendantes.

Des expériences sont *indépendantes* si la probabilité de tel ou tel résultat dans chacune des expériences ne dépend pas des résultats des autres. Par exemple, les expériences consistant à lancer une pièce de monnaie un certain nombre de fois sont indépendantes. Tirer des cartes d'un jeu de cartes c'est effectuer des expériences indépendantes. à condition toutefois que la carte tirée soit chaque fois remise dans le jeu et que les cartes soient battues; autrement, ces expériences ne sont pas indépendantes.

Des expériences indépendantes peuvent être effectuées dans les mêmes conditions ou dans des conditions différentes. Dans le premier cas la probabilité de l'événement A dans toutes les expériences est la même, dans le second cette probabilité change d'une expérience à l'autre. Pour le premier cas il y a lieu d'appliquer le *théorème particulier des expériences répétées* et pour le second le *théorème général.* Nous allons commencer par l'étude du théorème particulier qui est plus simple. Considérons d'abord un exemple.

E x e m p l e. Trois coups indépendants sont tirés sur une cible, la probabilité d'atteindre cette dernière lors de chaque coup est égale à p. Trouver la probabilité pour que deux de ces trois coups atteignent le but.

S o l u t i o n. Désignons par B_2 l'événement: « deux balles atteignent la cible ». Cet événement peut se réaliser de trois manières différentes:

1) le premier et le second coups bons, le troisième raté;
2) le premier et le troisième coups bons, le second raté;
3) le second et le troisième coups bons, le premier raté.

Par conséquent, l'événement B_2 peut être présenté comme la somme des produits des événements:

$$B_2 = A_1 A_2 \bar{A}_3 + A_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3,$$

où A_1, A_2, A_3 désignent les événements d'atteindre la cible lors du premier, du second, du troisième coup respectivement et $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ les événements de rater la cible la première, la seconde et la troisième fois.

Les trois variantes énumérées de l'événement B_2 étant incompatibles, et les événements composant chaque produit étant indépendants, les théorèmes d'addition et de multiplication donnent:

$$P(B_2) = pp(1-p) + p(1-p)p + (1-p)pp,$$

ou, en désignant $1 - p = q$,

$$P(B_2) = 3p^2 q.$$

D'une manière analogue, en examinant toutes les variantes possibles dans lesquelles un événement donné peut apparaître un certain nombre de fois on peut résoudre le problème général suivant.

Soient n expériences indépendantes, pouvant chacune soit réaliser l'événement A avec une probabilité p, soit ne pas le réaliser avec une probabilité $q = 1 - p$. Trouver la probabilité $P_{m,n}$ de réaliser l'événement A dans ces n expériences exactement m fois.

Considérons l'événement B_m correspondant à l'apparition de l'événement A lors de n expériences exactement m fois. Cet événement peut être réalisé de différentes manières. Nous allons décomposer l'événement B_m en une somme de produits d'événements consistant en l'apparition ou la non-apparition de l'événement A dans une expérience quelconque. Notons A_i l'apparition de l'événement A dans l'i-ème expérience, et $\bar{A}_i$ la non-apparition de l'événement A dans cette i-ème expérience.

Il est évident que chaque variante de réalisation de l'événement B_m (chaque terme de la somme) doit se composer de m apparitions de l'événement A et de $n - m$ non-apparitions, c'est-à-dire de m événements A et de $n - m$ événements $\bar{A}$ d'indices différents. Ainsi,

$$B_m = A_1 A_2 \ldots A_m \bar{A}_{m+1} \ldots \bar{A}_n + \ldots + A_1 \bar{A}_2 A_3 \ldots \bar{A}_{n-1} A_n + \\ + \ldots + \bar{A}_1 \bar{A}_2 \ldots \bar{A}_{n-m} A_{n-m+1} \ldots A_n,$$

dans chaque produit l'événement A devant figurer m fois et $\bar{A}$ $n - m$ fois.

Le nombre de combinaisons de ce genre est égal à C_n^m, c'est-à-dire au nombre de manières dont on peut choisir dans n expériences les m expériences dans lesquelles a eu lieu l'événement A. En vertu du théorème de multiplication des événements indépendants la probabilité de chacune de ces combinaisons est égale à $p^m q^{n-m}$. Les combinaisons étant mutuellement incompatibles, en vertu du théo-

rème d'addition, la probabilité de l'événement B_m vaut:

$$P_{m,n} = \underbrace{p^m q^{n-m} + \ldots + p^m q^{n-m}}_{C_n^m \text{ fois}} = C_n^m p^m q^{n-m}.$$

Ainsi, on peut formuler de la manière suivante le théorème particulier des expériences répétées.

Si on a n expériences indépendantes, dans chacune desquelles l'événement A apparaît avec une probabilité égale à p, la probabilité pour que l'événement A apparaisse exactement m fois est donnée par la formule suivante

$$P_{m,n} = C_n^m p^m q^{n-m}, \tag{4.1.1}$$

où $q = 1 - p$.

La formule (4.1.1) décrit la répartition des probabilités entre les valeurs possibles d'une certaine variable aléatoire, à savoir du nombre de réalisations de l'événement A dans n expériences.

Les probabilités $P_{m,n}$ ayant la forme des termes du développement du binôme $(q + p)^n$, la répartition des probabilités du type (4.1.1) est appelée *loi binomiale.*

4.2. Théorème général des expériences répétées

Le théorème particulier des expériences répétées n'est applicable qu'au cas où la probabilité de l'événement A dans toutes les expériences est la même. Plus fréquents sont les cas plus compliqués où les expériences ont lieu dans des conditions différentes et les probabilités de l'événement changent d'une expérience à l'autre. Si, par exemple, plusieurs coups sont tirés dans des conditions variées (disons, à des distances différentes), la probabilité d'atteindre le but peut notablement changer d'un coup à l'autre.

La méthode de calcul de la probabilité d'un nombre quelconque de réalisations de l'événement est dans ces conditions donnée par le *théorème général des expériences répétées.*

Supposons qu'on effectue n expériences indépendantes, dans chacune desquelles un certain événement A peut apparaître avec une probabilité p_i ou ne pas apparaître avec une probabilité $q_i = 1 - p_i$ $(i = 1, \ldots, n)$. Il y a lieu de trouver la probabilité $P_{m,n}$ pour que n expériences réalisent l'événement A exactement m fois.

Désignons comme auparavant par B_m l'événement suivant lequel l'événement élémentaire A apparaîtra m fois dans n expériences. Comme précédemment, B_m est la somme des produits d'événements élémentaires:

$$B_m = A_1 A_2 \ldots A_m \bar{A}_{m+1} \ldots \bar{A}_n + \ldots + A_1 \bar{A}_2 A_3 \ldots \bar{A}_{n-1} A_n + \\ + \bar{A}_1 \bar{A}_2 \ldots \bar{A}_{n-m} A_{n-m+1} \ldots A_n,$$

dans chacun des produits l'événement A entre m fois et l'événement $\bar{A}$ $n - m$ fois. Le nombre de ces combinaisons est comme auparavant égal à C_n^m, mais les combinaisons ne sont plus équiprobables.

En appliquant les théorèmes d'addition et de multiplication des probabilités pour les événements indépendants on obtient

$$P_{m,n} = p_1 p_2 \ldots p_m q_{m+1} \ldots q_n + \ldots$$
$$\ldots + p_1 q_2 p_3 \ldots q_{n-1} p_n + \ldots + q_1 q_2 \ldots q_{n-m} p_{n-m+1} \ldots p_n,$$

c'est-à-dire que la probabilité cherchée est égale à la somme de tous les produits possibles dans lesquels les probabilités p entrent m fois avec des indices différents et les probabilités q $n - m$ fois avec des indices différents.

Pour composer tous les produits possibles de m lettres p et de $n - m$ lettres q avec des indices différents nous employons le procédé suivant. Faisons le produit de n binômes

$$\varphi_n(z) = (q_1 + p_1 z)(q_2 + p_2 z) \ldots (q_n + p_n z)$$

ou dans l'écriture condensée

$$\varphi_n(z) = \prod_{i=1}^{n} (q_i + p_i z),$$

où z est un paramètre quelconque.

Proposons nous de trouver dans ce produit de binômes le coefficient de z^m. A cet effet, multiplions les binômes et réduisons les semblables. Il est clair que chaque terme contenant z^m aura pour coefficient le produit de m lettres p d'indices différents et de $n - m$ lettres q, et après la réduction des semblables le coefficient de z^m sera la somme de tous les produits possibles de ce type. Par conséquent, la méthode de recherche de ce coefficient est justement celle de calcul de la probabilité $P_{m,n}$ du problème d'expériences répétées.

La fonction $\varphi_n(z)$ dont le développement suivant les puissances du paramètre z a pour coefficients les probabilités $P_{m,n}$ s'appelle *fonction génératrice des probabilités* $P_{m,n}$ ou simplement *fonction génératrice*.

Utilisant la notion de fonction génératrice on peut formuler comme suit le théorème général sur les expériences répétées.

La probabilité pour que l'événement A apparaisse exactement m fois dans n expériences indépendantes est égale au coefficient de z^m dans l'expression de la fonction

$$\varphi_n(z) = \prod_{i=1}^{n} (q_i + p_i z),$$

où p_i est la probabilité d'apparition de l'événement A dans la i-ème expérience; $q_i = 1 - p_i$.

Au contraire du théorème particulier, la formulation ci-dessus du théorème général des expériences répétées ne donne pas la probabilité $P_{m,\,n}$ sous forme explicite. En principe, il est possible d'écrire cette expression, mais elle serait trop compliquée et nous ne la donnons pas ici. Néanmoins, sans utiliser cette forme explicite on peut quand même écrire le théorème général sur les expériences répétées comme une seule formule, soit:

$$\prod_{i=1}^{n} (q_i + p_i z) = \sum_{m=0}^{n} P_{m,\,n} z^m. \qquad (4.2.1)$$

Les deux membres de l'égalité (4.2.1) expriment une même fonction génératrice $\varphi_n(z)$ avec cette seule différence que dans le premier membre elle est écrite comme un monôme et dans le second comme un polynôme. En ouvrant les parenthèses dans le premier membre et en réduisant les termes semblables on obtient toutes les probabilités

$$P_{0,\,n},\ P_{1,\,n},\ \ldots,\ P_{n,\,n}$$

en tant que coefficients des puissances zéro, un, etc. de z.

Il est évident que le théorème particulier des expériences répétées découle du théorème général pour

$$p_1 = p_2 = \ldots = p_n = p,$$
$$q_1 = q_2 = \ldots = q_n = q.$$

Dans ce cas la fonction génératrice devient la n-ième puissance du binôme $(q + pz)$:

$$\varphi_n(z) = (q + pz)^n.$$

En développant cette expression suivant la formule du binôme on a:

$$(q + pz)^n = \sum_{m=0}^{n} C_n^m p^m q^{n-m} z^m,$$

d'où la formule (4.1.1).

Notons que tant dans le cas général que dans le cas particulier la somme de toutes les probabilités $P_{m,\,n}$ est égale à l'unité:

$$\sum_{m=0}^{n} P_{m,\,n} = 1. \qquad (4.2.2)$$

Ceci est dû avant tout à ce que les événements $B_0, B_1, \ldots, B_n$ forment un groupe complet d'événements incompatibles. Du point de vue formel on peut obtenir l'égalité (4.2.2) en posant dans la formule générale (4.2.1) $z = 1$.

Dans de nombreux cas pratiques en plus de la probabilité $P_{m,\,n}$ de réalisation de l'événement A exactement m fois, on a à envisager la probabilité d'au moins m réalisations de l'événement A.

Désignons par C_m l'événement correspondant à l'apparition de l'événement A au moins m fois, et par $R_{m,n}$ la probabilité de l'événement C_m. Il est évident que

$$C_m = B_m + B_{m+1} + \ldots + B_n,$$

d'où en vertu du théorème d'addition

$$R_{m,n} = P_{m,n} + P_{m+1,n} + \ldots + P_{n,n},$$

ou d'une façon plus condensée

$$R_{m,n} = \sum_{i=m}^{n} P_{i,n}. \qquad (4.2.3)$$

Il est souvent plus commode de calculer $R_{m,n}$ non pas directement par la formule (4.2.3) mais en passant à l'événement opposé:

$$R_{m,n} = 1 - \sum_{i=0}^{m-1} P_{i,n}. \qquad (4.2.4)$$

Exemple 1. Quatre coups indépendants sont tirés sur une même cible; les probabilités d'atteinte de ces coups sont respectivement:

$$p_1 = 0,1; \quad p_2 = 0,2; \quad p_3 = 0,3; \quad p_4 = 0,4.$$

Trouver les probabilités que zéro, un, deux, trois et quatre coups soient portés au but:

$$P_{0,4}; \quad P_{1,4}; \quad P_{2,4}; \quad P_{3,4}; \quad P_{4,4}.$$

Solution. Ecrivons la fonction génératrice:

$$\varphi_4(z) = \prod_{i=1}^{4} (q_i + p_i z) = (0,9 + 0,1z)(0,8 + 0,2z)(0,7 + 0,3z)(0,6 + 0,4z) =$$

$$= 0,302 + 0,440z + 0,215z^2 + 0,040z^3 + 0,002z^4.$$

d'où:

$$P_{0,4} = 0,302; \quad P_{1,4} = 0,440; \quad P_{2,4} = 0,215; \quad P_{3,4} = 0,040; \quad P_{4,4} = 0,002.$$

Exemple 2. Quatre coups indépendants sont tirés dans les mêmes conditions, la probabilité d'atteinte p étant la moyenne des probabilités p_1, p_2, p_3, p_4 de l'exemple précédent:

$$p = \frac{1}{4}(p_1 + p_2 + p_3 + p_4) = 0,25.$$

Trouver les probabilités

$$P_{0,4}; \quad P_{1,4}; \quad P_{2,4}; \quad P_{3,4}; \quad P_{4,4}.$$

Solution. En vertu de la formule (4.1.1) on a:

$$P_{0,4} = q^4 = 0,316;$$
$$P_{1,4} = C_4^1 pq^3 = 0,421;$$
$$P_{2,4} = C_4^2 p^2q^2 = 0,211;$$
$$P_{3,4} = C_4^3 p^3q = 0,047;$$
$$P_{4,4} = p^4 = 0,004.$$

E x e m p l e 3. La communication est maintenue avec 5 stations. La liaison est parfois interrompue à cause des conditions météorologiques. Les stations étant suffisamment éloignées les unes des autres, les arrêts de liaison sont indépendants et se produisent avec une probabilité $p = 0{,}2$. Trouver la probabilité pour qu'à un instant donné la liaison soit maintenue avec au plus deux stations.

S o l u t i o n. L'événement dont il est question consiste en ce qu'il n'y aura pas de communication avec au moins trois stations. En vertu de la formule (4.2.3) on a:

$$R_{3,5} = P_{3,5} + P_{4,5} + P_{5,5} = C_5^3 \cdot 0{,}2^3 \cdot 0{,}8^2 + \\ + C_5^4 \cdot 0{,}2^4 \cdot 0{,}8 + 0{,}2^5 = 0{,}0512 + 0{,}0064 + 0{,}0003 = 0{,}0579$$

E x e m p l e 4. Un système de stations radars poursuit un groupe de 10 objectifs. Chaque objectif peut être perdu indépendamment des autres avec une probabilité égale à 0,1. Trouver la probabilité de perdre au moins un objectif.

S o l u t i o n. La probabilité $R_{1,10}$ pour qu'au moins un objectif soit perdu peut être obtenue par la formule:

$$R_{1,10} = P_{1,10} + P_{2,10} + \ldots + P_{10,10},$$

mais il est bien plus simple d'utiliser la probabilité de l'événement contraire, consistant en ce qu'aucun objectif n'est perdu, et de la retrancher de l'unité:

$$R_{1,10} = 1 - P_{0,10} = 1 - 0{,}9^{10} \approx 0{,}65.$$

E x e m p l e 5. Un appareil se compose de 8 éléments semblables, mais il ne peut fonctionner que si au moins six d'entre eux fonctionnent. Durant le temps t chaque élément peut tomber en panne indépendamment des autres avec une probabilité égale à 0,2. Trouver la probabilité de la défaillance de l'appareil durant le temps t.

S o l u t i o n. Pour que l'appareil soit en panne il faut qu'au moins deux des huit éléments cessent de fonctionner. La formule (4.2.4) donne:

$$R_{2,8} = 1 - (P_{0,8} + P_{1,8}) = 1 - (0{,}8^8 + C_8^1 \cdot 0{,}2 \cdot 0{,}8^7) \approx 0{,}497.$$

E x e m p l e 6. Un appareil se compose de cinq éléments; la fiabilité (probabilité de fonctionnement sans défaillance durant un temps t) du premier élément est $p_1 = 0{,}9$, du second $p_2 = 0{,}95$, du troisième $p_3 = 0{,}8$, du quatrième $p_4 = 0{,}85$ et du cinquième $p_5 = 0{,}91$. Si aucun élément n'est en panne la probabilité de fonctionnement sans défaillance de l'appareil est égale à 1; si un des cinq éléments est en panne, cette probabilité est 0,7; si plus de deux éléments sont en panne l'appareil ne peut fonctionner.

Trouver la probabilité pour que l'appareil puisse effectuer le travail qui lui est assigné.

S o l u t i o n. Pour résoudre le problème on utilise la formule de la probabilité totale avec les hypothèses suivantes:

H_0 aucun élément n'est en panne;

H_1 un élément est en panne.

Pour trouver les probabilités des hypothèses on utilise le théorème général des expériences répétées, et on écrit la fonction génératrice comme suit:

$$\varphi_5(z) = \prod_{i=1}^{5} (q_i + p_i z) = \\ = (0{,}9 + 0{,}1z)(0{,}95 + 0{,}05z)(0{,}8 + 0{,}2z)(0{,}85 + 0{,}15z)(0{,}91 + 0{,}09z).$$

Il n'est pas nécessaire ici de calculer tous les termes du développement de la fonction génératrice suivant les puissances de z; il suffit de trouver les

deux premiers termes, notamment, le terme constant et le terme correspondant à la première puissance de z. On a:

$$\varphi_5(z) = 0{,}9\cdot 0{,}95\cdot 0{,}8\cdot 0{,}85\cdot 0{,}91 + (0{,}1\cdot 0{,}95\cdot 0{,}8\cdot 0{,}85\cdot 0{,}91 + \\ + 0{,}9\cdot 0{,}05\cdot 0{,}8\cdot 0{,}85\cdot 0{,}91 + 0{,}9\cdot 0{,}95\cdot 0{,}2\cdot 0{,}85\cdot 0{,}91 + \\ + 0{,}9\cdot 0{,}95\cdot 0{,}8\cdot 0{,}15\cdot 0{,}91 + 0{,}9\cdot 0{,}95\cdot 0{,}8\cdot 0{,}85\cdot 0{,}09)\, z + \ldots = \\ = 0{,}530 + 0{,}364z + \ldots$$

D'où

$$P(H_0) = 0{,}530; \quad P(H_1) = 0{,}364.$$

En vertu du théorème de la probabilité totale, la probabilité de l'événement A que l'appareil effectuerait le travail qui lui est assigné est égale à:

$$P(A) = 0{,}530\cdot 1 + 0{,}364\cdot 0{,}7 = 0{,}784.$$

CHAPITRE 5

VARIABLES ALÉATOIRES ET LOIS DE RÉPARTITION

5.1. Suite d'une répartition. Polygone d'une répartition

Dans la partie du cours consacrée aux notions fondamentales de la théorie des probabilités nous avons introduit la notion très importante de *variable aléatoire*. Nous développons ici cette notion et indiquons les méthodes permettant de décrire et de caractériser les variables aléatoires.

Rappelons qu'on appelle variable aléatoire une grandeur pouvant prendre lors d'une expérience telle ou telle valeur, inconnue d'avance. On distingue les variables aléatoires discontinues (discrètes) et les variables aléatoires continues. L'ensemble des valeurs possibles des variables discrètes est dénombrable. Les valeurs possibles des variables continues ne peuvent être dénombrées et remplissent d'une manière continue un certain intervalle.

A titre d'exemple de variables aléatoires discrètes on peut citer:
1) le nombre d'apparitions de face lorsqu'on jette une pièce trois fois de suite (les valeurs possibles sont 0, 1, 2, 3);
2) la fréquence d'apparition de face dans l'expérience précédente (les valeurs possibles sont 0, 1/3, 2/3, 1);
3) le nombre d'éléments défectueux dans un appareil formé de cinq éléments (les valeurs possibles sont 0, 1, 2, 3, 4, 5);
4) le nombre de coups avant la première atteinte de la cible (les valeurs possibles sont 1, 2, 3, . . ., n, . . .);
5) le nombre d'étudiants ayant passé leurs examens (les valeurs possibles sont 0, 1, 2, . . ., N, où N est le nombre total d'élèves).

A titre d'exemple de variables aléatoires continues on peut citer:
1) l'abscisse (l'ordonnée) d'un point en mouvement;
2) la distance du point d'atteinte au centre de la cible;
3) l'erreur d'un altimètre;
4) le temps de fonctionnement sans défaillance d'un tube électronique;
5) la durée de vie d'un satellite artificiel de la Terre.

Convenons de désigner les variables aléatoires par les majuscules et leurs valeurs possibles par les minuscules correspondantes. Par

exemple, X est le nombre d'atteintes lors de trois coups, les valeurs possibles étant $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4=3$.

Considérons maintenant la variable aléatoire X dont les valeurs possibles sont $x_1, x_2, \ldots, x_n$. La variable X peut prendre chacune d'elles avec une certaine probabilité. Le fait de prendre pour X au cours d'une expérience l'une de ces valeurs est la réalisation d'un des événements formant le système complet d'événements incompatibles:

$$\left.\begin{array}{l} X = x_1, \\ X = x_2, \\ \ldots\ldots, \\ X = x_n. \end{array}\right\} \qquad (5.1.1)$$

Désignons par p avec l'indice correspondant les probabilités de ces événements:

$$P(X = x_1) = p_1; \quad P(X = x_2) = p_2; \ldots; P(X = x_n) = p_n.$$

Les événements incompatibles (5.1.1) formant un système complet, on a:

$$\sum_{i=1}^{n} p_i = 1,$$

c'est-à-dire que la somme des probabilités de toutes les valeurs possibles d'une variable aléatoire est égale à l'unité. Cette probabilité totale est d'une certaine façon répartie suivant les différentes valeurs. Donner cette répartition, c'est-à-dire indiquer exactement la probabilité de chacun des événements (5.1.1), c'est décrire une variable aléatoire du point de vue probabiliste. On établit ainsi la *loi de répartition* de la variable aléatoire.

On appelle *loi de répartition* d'une variable aléatoire toute relation établissant une correspondance entre les valeurs possibles de cette variable aléatoire et leurs probabilités. On dira alors que la variable aléatoire *suit* la loi de répartition en question.

Voyons comment on peut donner la loi de répartition d'une variable aléatoire discrète X. La forme la plus simple est un tableau où se trouvent inscrites les valeurs possibles de X et les probabilités correspondantes:

x_i	x_1	x_2	$\ldots$	x_n
p_i	p_1	p_2	$\ldots$	p_n

Nous appellerons ce tableau *suite de répartition de X*.

Pour plus de clarté on représente souvent un tableau de répartition sous forme d'un graphe, en portant en abscisses les valeurs possibles de la variable aléatoire, et en ordonnées les probabilités

de ces valeurs. Ce qu'on obtient est appelé *diagramme en bâtons*. En réunissant les extrémités des barres verticales représentant les probabilités par des droites, on obtient un *polygone de répartition* (fig. 5.1.1), qui, tout comme le tableau de répartition caractérise complètement la variable aléatoire qu'il définit; c'est là une des formes possibles de la loi de répartition.

Il est parfois commode de se servir de l'interprétation mécanique d'une loi de répartition. Supposons qu'une certaine masse égale à l'unité soit répartie suivant l'axe des abscisses de telle sorte qu'en

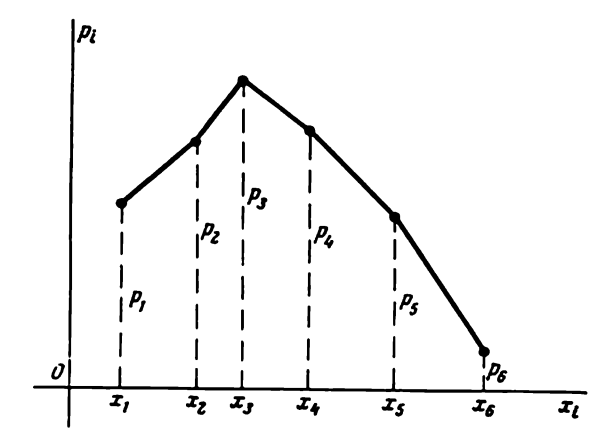


Fig. 5.1.1

n points $x_1, x_2, \ldots, x_n$ se trouvent concentrées les masses $p_1, p_2, \ldots$ $\ldots, p_n$. La suite de répartition est alors interprétée comme un système de points matériels, de masses déterminées, disposés sur l'axe des abscisses.

Considérons quelques exemples de variables aléatoires discrètes et leurs lois de répartition.

E x e m p l e 1. Soit une expérience pouvant donner lieu à l'événement A. La probabilité de réalisation de l'événement A est égale à 0,3. On considère la variable aléatoire X représentant le nombre de réalisations de l'événement A dans l'expérience en question (c'est-à-dire la variable indicatrice de l'événement A qui prend la valeur 1 s'il se réalise et la valeur 0 dans le cas contraire). Etablir le tableau de répartition et construire le polygone de répartition de X.

S o l u t i o n. La variable X a deux valeurs: 0 et 1. Le tableau de répartition de X est:

x_i	0	1
p_i	0,7	0,3

Le polygone de répartition est représenté sur la figure 5.1.2.

E x e m p l e 2. Trois coups sont tirés sur une cible. La probabilité d'atteindre la cible avec un coup est égale à 0,4. Pour chaque coup réussi le tireur gagne 5 points. Etablir le tableau de répartition du nombre de points gagnés.

S o l u t i o n. Désignons par X le nombre de points gagnés. Les valeurs possibles de X sont $x_1 = 0$, $x_2 = 5$; $x_3 = 10$, $x_4 = 15$.

Les probabilités de ces valeurs sont données par le théorème des expériences répétées:

$$p_1 = 0{,}6^3 = 0{,}216; \qquad p_2 = C_3^1 \cdot 0{,}4 \cdot 0{,}6^2 = 0{,}432;$$

$$p_3 = C_3^2 \cdot 0{,}4^2 \cdot 0{,}6 = 0{,}288; \qquad p_4 = 0{,}4^3 = 0{,}064.$$

Le tableau de répartition de X sera:

x_i	0	5	10	15
p_i	0,216	0,432	0,288	0,064

Le polygone de répartition est donné sur la figure 5.1.3.

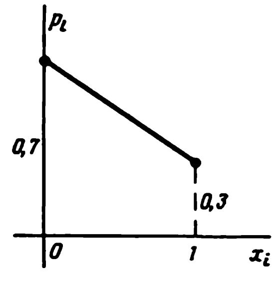


Fig. 5.1.2

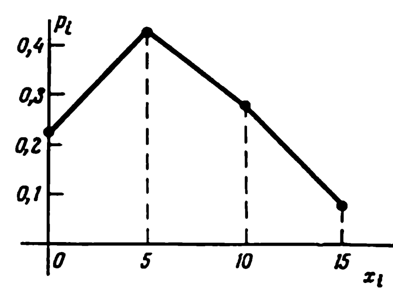


Fig. 5.1.3

E x e m p l e 3. La probabilité de réalisation de l'événement A dans une expérience est égale à p. On répète cette expérience (les reprises étant indépendantes) jusqu'à la première réalisation de l'événement A, après quoi les expériences sont arrêtées. Notons X le nombre d'expériences effectuées. Trouver le tableau de répartition de X.

S o l u t i o n. Les valeurs possibles de X sont 1, 2, 3, . . . (théoriquement elles ne sont limitées en rien). Pour que X prenne la valeur 1 il faut que l'événement A ait lieu dans la première expérience; la probabilité de cette éventualité est égale à p. Pour que X prenne la valeur 2, il faut que dans la première expérience l'événement A ne se réalise pas mais se produise dans la seconde; ceci a une probabilité qp, où $q = 1 - p$, etc. Le tableau de répartition de X est de la forme:

x_i	1	2	3	...	i	...
p_i	p	pq	pq^2	...	pq^{i-1}	...

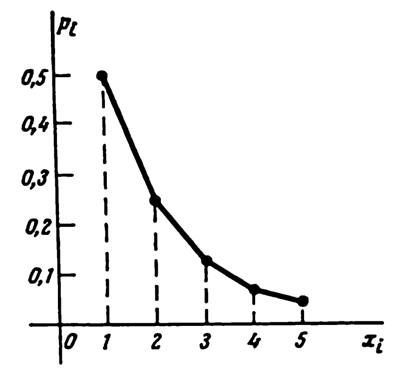


Fig. 5.1.4

Les cinq premières ordonnées du polygone de répartition pour le cas $p = q = 0{,}5$ sont montrées sur la figure 5.1.4.

E x e m p l e 4. Un tireur tire sur une cible jusqu'à la première atteinte ayant à sa disposition 4 cartouches. La probabilité d'atteindre le but lors de

chaque coup est égale à 0,6. Trouver le tableau de répartition des cartouches non dépensées.

S o l u t i o n. La variable aléatoire X représentant le nombre de cartouches non dépensées peut prendre quatre valeurs possibles: 0, 1, 2 et 3 avec les probabilités:

$$p_0 = 0{,}4^3 = 0{,}064\,;$$
$$p_1 = 0{,}4^2 \cdot 0{,}6 = 0{,}096\,;$$
$$p_2 = 0{,}4 \cdot 0{,}6 = 0{,}240\,;$$
$$p_3 = 0{,}600.$$

Le tableau de répartition de X est:

x_i	0	1	2	3
p_i	0,064	0,096	0,240	0,600

Le polygone de répartition est donné sur la figure 5.1.5.

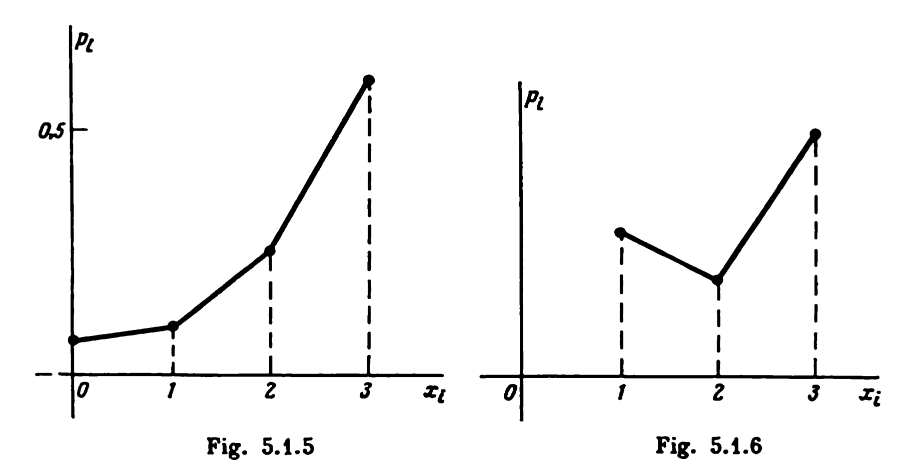


Fig. 5.1.5 Fig. 5.1.6

E x e m p l e 5. Un appareil technique peut être utilisé dans différentes conditions. Il exige d'être réglé de temps à autre suivant les conditions d'utilisation. Choisi pour un certain service, il peut se trouver au hasard dans un régime favorable ou défavorable. Dans le premier cas il peut être utilisé trois fois de suite sans réglage, puis il doit être réglé. Dans le second cas il doit être réglé après la première utilisation. La probabilité pour que l'appareil soit dans un régime favorable est égale à 0,7, celle d'un régime défavorable vaut 0,3. On considère la variable aléatoire X, le nombre d'utilisations de l'appareil avant le réglage; trouver le tableau de répartition de X.

S o l u t i o n. X peut prendre trois valeurs: 1, 2 et 3. Pour que $X = 1$ il faut que dès la première utilisation l'appareil se trouve dans un régime défavorable, la probabilité de cette éventualité est $p_1 = 0{,}3$. Pour que $X = 2$ il faut que lors de la première utilisation l'appareil se trouve dans un régime favorable et lors de la seconde dans un régime défavorable, la probabilité correspondante est $p_2 = 0{,}7 \cdot 0{,}3 = 0{,}21$. Pour que $X = 3$ il faut que lors des deux premières utilisations l'appareil soit dans un régime favorable (après la troisième utilisation le réglage est obligatoire); la probabilité correspondante est $p_3 = 0{,}7^2 = 0{,}49$.

Le tableau de répartition de X est:

x_i	1	2	3
p_i	0,30	0,21	0,49

Le polygone de répartition est donné sur la figure 5.1.6.

5.2. Fonction de répartition

Dans le paragraphe précédent nous avons introduit le tableau de répartition (suite ordonnée) en tant que caractéristique complète (loi de répartition) d'une variable aléatoire discrète. Cependant cette caractéristique n'est pas universelle; elle n'a de sens que pour des variables aléatoires discrètes. Il est facile de voir que pour une variable aléatoire continue cette caractéristique ne peut être utilisée. En effet, une variable aléatoire continue a une infinité de valeurs possibles remplissant d'une manière continue un certain intervalle (ensemble dit non dénombrable). Il est donc impossible de dresser un tableau où toutes les valeurs d'une telle variable aléatoire sont énumérées. De plus, comme nous allons le voir ultérieurement, chaque valeur particulière de la variable aléatoire continue n'a pas de probabilité différente de zéro. Cependant les différents domaines des valeurs possibles d'une variable aléatoire ne sont pas équiprobables, et pour une variable continue il existe une certaine répartition des probabilités, bien qu'il faille entendre ceci d'une manière quelque peu différente du cas d'une variable discrète.

Pour caractériser quantitativement cette répartition il est commode d'utiliser non pas la probabilité de l'événement $X = x$, mais la probabilité de l'événement $X < x$ où x est une certaine valeur courante. La probabilité de cet événement dépend évidemment de x, elle en est une certaine fonction. Cette fonction est appelée *fonction de répartition* de la variable aléatoire X. Il est d'usage de la désigner par $F(x)$:

$$F(x) = P(X < x). \tag{5.2.1}$$

La fonction de répartition est la caractéristique la plus universelle d'une variable aléatoire. Elle existe pour toutes les variables aléatoires, tant discrètes que continues. La fonction de répartition caractérise complètement une variable aléatoire du point de vue probabiliste, c'est-à-dire se trouve être une des formes de la loi de répartition.

Mentionnons quelques propriétés générales de la fonction de répartition.

1. La fonction de répartition $F(x)$ est une fonction non décroissante de son argument, c'est-à-dire que pour $x_2 > x_1$ on a $F(x_2) \geqslant F(x_1)$.

2. Lorsque l'argument $x \to -\infty$, la fonction de répartition est égale à zéro:

$$F(-\infty) = 0.$$

3. Lorsque l'argument $x \to +\infty$, la fonction de répartition est égale à l'unité:

$$F(+\infty) = 1.$$

Sans donner de démonstration rigoureuse de ces propriétés nous allons les rendre claires par une interprétation géométrique. A cet effet considérons la variable aléatoire X comme un *point aléatoire* X

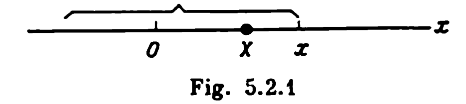


Fig. 5.2.1

sur l'axe Ox (fig. 5.2.1) pouvant dans l'expérience se situer dans une position quelconque. La fonction de répartition $F(x)$ est alors la probabilité pour que le point aléatoire X se trouve *à gauche* du point x.

Faisons augmenter l'abscisse x, c'est-à-dire déplaçons le point x à droite. On comprend aisément que la probabilité pour le point

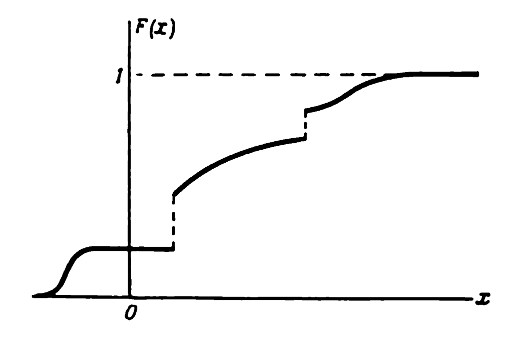


Fig. 5.2.2

aléatoire X de se trouver à gauche de x ne peut alors diminuer; par conséquent, la fonction de répartition $F(x)$ ne peut décroître quand x croît.

Pour voir que $F(-\infty) = 0$ on déplace indéfiniment le point x vers la gauche. Il devient impossible pour le point X de se trouver à la limite à gauche de x; il est naturel de supposer que la probabilité de cet événement tend vers zéro, c'est-à-dire que $F(-\infty) = 0$.

De même, en déplaçant indéfiniment le point x à droite on voit que $F(+\infty) = 1$, car l'événement $X < x$ devient à la limite certain.

Le graphique de la fonction de répartition $F(x)$ est dans le cas général une courbe non décroissante (fig. 5.2.2) variant de 0 à 1, avec des discontinuités éventuelles en certains points.

Connaissant le tableau de répartition d'une variable aléatoire on peut facilement trouver sa fonction de répartition, En effet,

$$F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i),$$

où l'inégalité $x_i < x$ sous le signe de la somme indique que la sommation est étendue à toutes les x_i inférieures à x.

Lorsque x passe par une des valeurs possibles de la variable discrète X, la fonction varie par saut, la grandeur du saut étant égale à la probabilité de cette valeur possible.

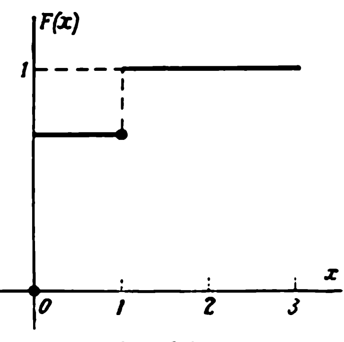


Fig. 5.2.3

E x e m p l e 1. On effectue une expérience pouvant réaliser ou ne pas réaliser l'événement A. La probabilité de réalisation de l'événement A est égale à 0.3. On considère la variable aléatoire X, le nombre de réalisations de l'événement A dans l'expérience (variable indicatrice de l'événement A). Trouver sa fonction de répartition.

S o l u t i o n. Le tableau de répartition de X est:

x_i	0	1
p_i	0,7	0,3

Construisons la fonction de répartition de X:

1) pour $x \leqslant 0$

$$F(x) = P(X < x) = 0;$$

2) pour $0 < x \leqslant 1$

$$F(x) = P(X < x) = P(X = 0) = 0,7$$

3) pour $x > 1$

$$F(x) = P(X < x) = P(X = 0) + P(X = 1) = 1.$$

La figure 5.2.3 montre le graphique de la fonction de répartition. Aux points de discontinuité la fonction $F(x)$ prend les valeurs notées sur la figure par des points (la fonction est continue à gauche).

E x e m p l e 2. Dans les conditions de l'exemple précédent on effectue 4 expériences indépendantes. Déterminer la fonction de répartition du nombre de réalisations de l'événement A.

S o l u t i o n. Désignons par X le nombre de réalisations de l'événement A dans quatre expériences. Voici son tableau de répartition:

x_i	0	1	2	3	4
p_i	0,2401	0,4116	0,2646	0,0756	0,0081

Construisons la fonction de répartition de X:
1) pour $x \leqslant 0$ $F(x) = 0$;
2) pour $0 < x \leqslant 1$ $F(x) = 0{,}2401$;
3) pour $1 < x \leqslant 2$ $F(x) = 0{,}6517$;
4) pour $2 < x \leqslant 3$ $F(x) = 0{,}9163$;
5) pour $3 < x \leqslant 4$ $F(x) = 0{,}9919$;
6) pour $x > 4$ $F(x) = 1$.

Sur la figure 5.2.4 on peut voir le graphique de la fonction de répartition $F(x)$.

La fonction de répartition d'une variable discrète quelconque est toujours une fonction discontinue en escalier dont les sauts se produisent aux points qui sont les valeurs possibles de cette variable

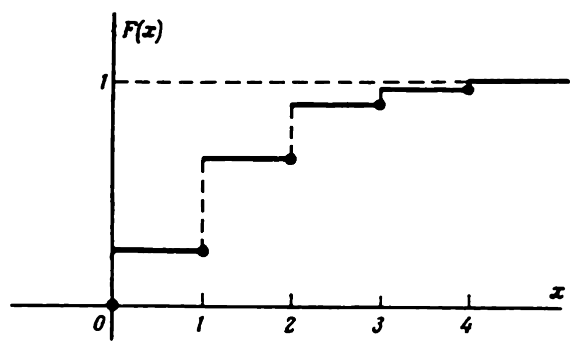


Fig. 5.2.4

aléatoire et sont égaux aux probabilités de ces valeurs. La somme de tous les sauts (échelons) de la fonction $F(x)$ est égale à l'unité.

Au fur et à mesure que le nombre de valeurs possibles de la variable aléatoire augmente et que diminuent les intervalles entre elles, le nombre d'échelons devient plus important et les échelons plus petits; la courbe en escalier devient plus continue (fig. 5.2.5); la variable aléatoire tend peu à peu vers une variable continue, et sa fonction de répartition vers une fonction continue (fig. 5.2.6).

Généralement, dans les applications pratiques la fonction de répartition d'une variable aléatoire continue est une fonction continue en tous les points comme on peut le voir sur la figure 5.2.6. On peut cependant trouver des exemples de variables aléatoires dont les valeurs possibles remplissent d'une manière continue un certain intervalle, mais pour lesquelles la fonction de répartition n'est pas partout continue mais admet des discontinuités (fig. 5.2.7). Ces variables aléatoires sont dites *mixtes*. On peut citer à titre d'exemple la surface de destruction d'une cible par une bombe dont le rayon destructif est égal à R (fig. 5.2.8). Les valeurs de cette variable aléatoire remplissent l'intervalle compris entre 0 et πR^2, mais les valeurs limites 0 et πR^2 de cet intervalle correspondant aux

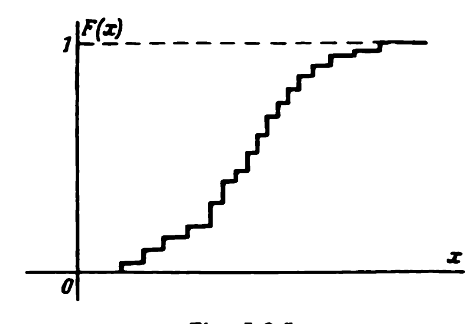


Fig. 5.2.5

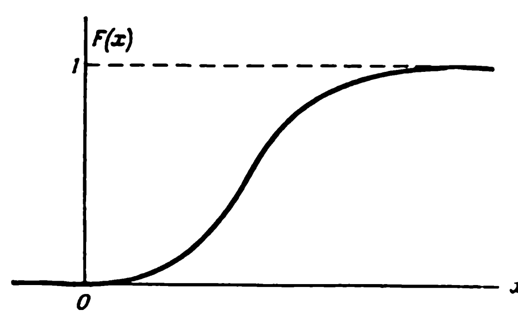


Fig. 5.2.6

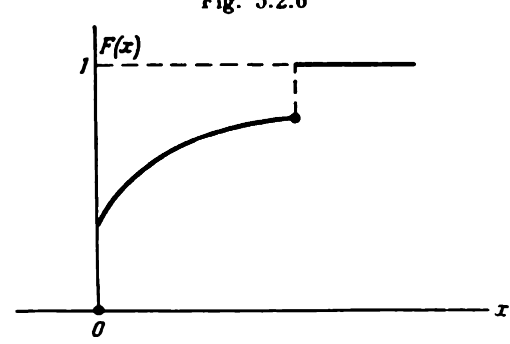


Fig. 5.2.7

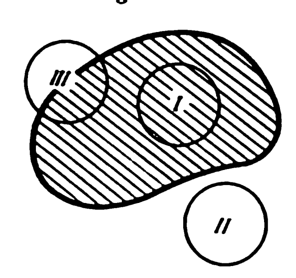


Fig. 5.2.8

positions I et II de la bombe ont une certaine probabilité finie, à ces valeurs correspondent des sauts de la fonction de répartition, alors que pour les valeurs intermédiaires (position du type III) la fonction de répartition est continue. Pour autre exemple de la variable aléatoire mixte on peut citer le temps T de fonctionnement sans défaillance d'un appareil soumis aux essais durant le temps t. La fonction de répartition de cette variable aléatoire est continue partout, à l'exception du point t.

5.3. Probabilité pour une variable aléatoire de tomber dans un intervalle donné

Dans les applications pratiques souvent il y a lieu de calculer la probabilité pour une variable aléatoire de prendre une valeur comprise dans certaines limites, entre α et β par exemple. Nous appellerons cet événement « l'appartenance de la variable aléatoire X à l'intervalle compris entre α et β ».

Pour fixer les idées convenons que l'extrémité gauche de l'intervalle (α, β) lui appartient et l'extrémité droite ne lui appartient pas. Alors, exiger que la variable X tombe dans l'intervalle en question c'est assurer que l'inégalité suivante soit vérifiée:

$$\alpha \leqslant X < \beta.$$

Exprimons la probabilité de cet événement à l'aide de la fonction de répartition de X. A cet effet considérons les trois événements suivants:

l'événement A consistant en ce que $X < \beta$;
l'événement B consistant en ce que $X < \alpha$;
l'événement C consistant en ce que $\alpha \leqslant X < \beta$.

Comme $A = B + C$, on a en vertu du théorème d'addition des probabilités:

$$P(X < \beta) = P(X < \alpha) + P(\alpha \leqslant X < \beta)$$

ou

$$F(\beta) = F(\alpha) + P(\alpha \leqslant X < \beta),$$

d'où

$$P(\alpha \leqslant X < \beta) = F(\beta) - F(\alpha), \tag{5.3.1}$$

c'est-à-dire que la *probabilité pour la variable aléatoire de tomber dans un certain intervalle est égale à l'accroissement de la fonction de répartition sur cet intervalle.*

Si on diminue indéfiniment le domaine (α, β) en supposant que $\beta \to \alpha$, au lieu de la probabilité d'un certain intervalle on obtient à la limite la probabilité pour la variable de prendre une valeur

particulière α:

$$P(X=\alpha)=\lim_{\beta\to\alpha} P(\alpha \leqslant X < \beta)=\lim_{\beta\to\alpha}[F(\beta)-F(\alpha)]. \quad (5.3.2)$$

La valeur de cette limite dépend de ce que la fonction $F(x)$ est continue au point $x = \alpha$ ou admet en ce point une discontinuité. Si au point α la fonction $F(x)$ a une discontinuité, la limite de (5.3.2) est égale à la valeur du saut de la fonction $F(x)$ en ce point. Si au contraire la fonction $F(x)$ est continue au point α, cette limite est *égale à zéro*.

Ultérieurement nous dirons que des variables aléatoires sont continues seulement si leurs fonctions de répartition le sont partout. On peut donc formuler la propriété suivante:

La probabilité d'une valeur quelconque d'une variable continue est égale à zéro.

Nous allons nous arrêter un peu plus en détail sur cette propriété. Dans ce cours nous avons déjà rencontré des événements dont la probabilité était égale à zéro, il s'agissait d'événements impossibles. On voit maintenant que non seulement les événements impossibles, mais également les événements possibles peuvent avoir une probabilité nulle. En effet, l'événement $X = \alpha$ est possible; cependant la probabilité de cet événement est égale à zéro. De tels événements, possibles mais avec une probabilité nulle, ne peuvent apparaître que dans des expériences ne se réduisant pas à un système de cas.

La notion d'événement possible mais avec une probabilité nulle peut à première vue paraître paradoxale. En réalité cette notion n'est pas plus paradoxale que celle d'un corps ayant une certaine masse. alors qu'aucun point à l'intérieur du corps n'a de masse finie. Si petit que soit un volume, sa masse est finie; cette masse tend vers zéro lorsque le volume diminue infiniment et à la limite est nulle pour un point. De même, pour une répartition continue la probabilité pour un point aléatoire de tomber dans un intervalle aussi petit que l'on veut peut être différente de zéro, alors que la probabilité de se trouver en un point déterminé est exactement nulle. Soit une expérience dans laquelle la variable aléatoire continue X doit prendre une de ces valeurs possibles. Avant l'expérience la probabilité de chacune de ces valeurs est nulle; néanmoins dans l'expérience la variable aléatoire X prendra obligatoirement l'une de ses valeurs possibles, c'est-à-dire que de toute évidence va se réaliser l'un des événements de probabilité nulle.

Si la probabilité de l'événement $X = \alpha$ est nulle, ceci ne veut pas dire que cet événement n'apparaîtra pas, donc que la fréquence de cet événement est nulle. On a déjà vu que pour un grand nombre d'expériences la fréquence d'un événement n'est pas égale, mais tend seulement vers sa probabilité. Le fait que la probabilité de l'événement $X = \alpha$ est nulle veut simplement dire que pour un nombre

d'expériences indéfiniment grand cet événement se produira aussi rarement que l'on veut.

Si un événement A a une probabilité nulle d'être réalisé dans l'expérience en question, l'événement contraire $\bar{A}$ a une probabilité égale à l'unité, néanmoins il n'est pas certain. Pour une variable aléatoire continue X pour α quelconque la probabilité de l'événement $X \neq \alpha$ a une probabilité égale à l'unité, cependant cet événement n'est pas certain. Si le nombre d'expériences augmente indéfiniment cet événement aura lieu presque toujours, mais pas toujours.

Dans le paragraphe 5.1 nous avons interprété une variable aléatoire discrète comme une masse unité répartie entre certains points isolés de l'axe des abscisses. Dans le cas d'une variable aléatoire continue l'interprétation mécanique revient à la répartition d'une masse unité non pas en des points isolés, mais continûment sur l'axe des abscisses, la masse d'aucun des points n'étant finie.

5.4. Densité de probabilité

Soit une variable aléatoire continue X donnée par sa fonction de répartition $F(x)$ que nous supposerons continue et dérivable. Calculons la probabilité pour cette variable aléatoire de se trouver dans l'intervalle compris entre x et $x + \Delta x$:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x),$$

donc l'accroissement de la fonction de répartition sur cet intervalle. Considérons le rapport de cette probabilité à la longueur de l'intervalle, c'est-à-dire la probabilité moyenne par unité de longueur de l'intervalle et faisons tendre Δx vers zéro. A la limite on obtiendra la dérivée de la fonction de répartition:

$$\lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x). \qquad (5.4.1)$$

Introduisons la désignation

$$f(x) = F'(x). \qquad (5.4.2)$$

La fonction $f(x)$ — dérivée de la fonction de répartition — caractérise la densité de la répartition des valeurs de la variable aléatoire en un point donné. Cette fonction est appelée *densité de répartition des probabilités* ou plus brièvement *densité de probabilité* de la variable aléatoire continue X.

Les termes « densité de répartition », « densité de probabilité » deviennent particulièrement clairs si l'on utilise l'interprétation mécanique d'une répartition; dans cette interprétation la fonction $f(x)$ caractérise littéralement la densité de la répartition des masses sur l'axe des abscisses (« densité linéaire »). On a montré sur la figure 5.4.1 une *courbe de densité de probabilité*.

La densité de probabilité tout comme la fonction de répartition est une des formes de la loi de répartition. Au contraire de la fonction de répartition, cette forme n'est pas universelle, car elle n'a de sens que pour des variables aléatoires continues.

Considérons une variable aléatoire continue X de densité de probabilité $f(x)$ et le domaine élémentaire dx adjacent au point x (fig. 5.4.2). La probabilité pour la variable aléatoire X de se trouver dans cet intervalle élémentaire (à des infiniment petits d'ordre élevé

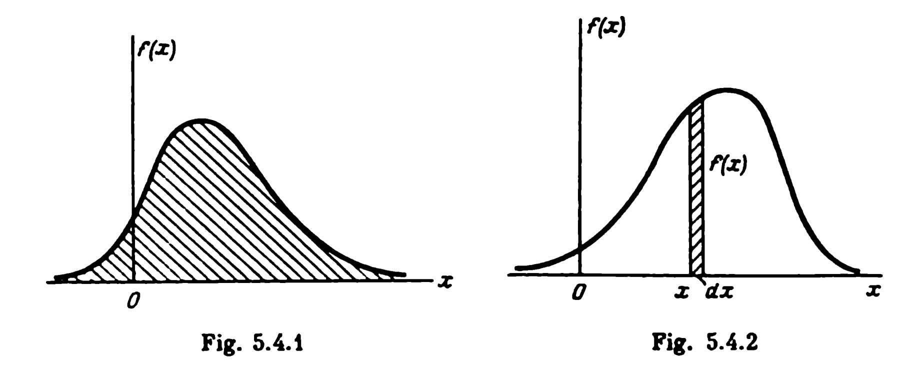


Fig. 5.4.1 Fig. 5.4.2

près) est égale à $f(x)\,dx$. La grandeur $f(x)\,dx$ est appelée *élément de probabilité*. Du point de vue géométrique c'est l'aire du rectangle élémentaire ayant pour base l'intervalle dx (fig. 5.4.2).

Exprimons la probabilité pour la variable aléatoire X de se trouver dans l'intervalle (α, β) (fig. 5.4.3) en fonction de la densité de probabilité. Il est évident que la probabilité cherchée est égale à la somme des éléments de probabilité sur tout cet intervalle, c'est-à-dire à l'intégrale suivante:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx \,^{*}). \qquad (5.4.3)$$

Du point de vue géométrique la probabilité pour la variable aléatoire X de se trouver dans l'intervalle (α, β) est égale à l'aire comprise entre la courbe de densité et l'axe des abscisses et limitée par les ordonnées correspondant aux extrémités de l'intervalle (α, β) (fig. 5.4.3).

La formule (5.4.2) exprime une densité de probabilité à l'aide d'une fonction de répartition. Soit maintenant le problème inverse,

* La probabilité d'une valeur quelconque d'une variable aléatoire continue étant égale à zéro, on peut considérer ici que l'extrémité gauche n'appartient pas à l'intervalle (α, β), c'est-à-dire en omettant le signe d'égalité dans $\alpha \leqslant X < \beta$.

c'est-à-dire qu'il y a lieu de trouver une fonction de répartition en fonction d'une densité. Par définition on a:

$$F(x) = P(X < x) = P(-\infty < X < x),$$

alors en vertu de la formule (5.4.3):

$$F(x) = \int_{-\infty}^{x} f(x)\,dx. \qquad (5.4.4)$$

Du point de vue géométrique, $F(x)$ n'est rien d'autre que l'aire comprise entre la courbe de densité et l'axe des abscisses et se trouvant à gauche du point x (fig. 5.4.4).

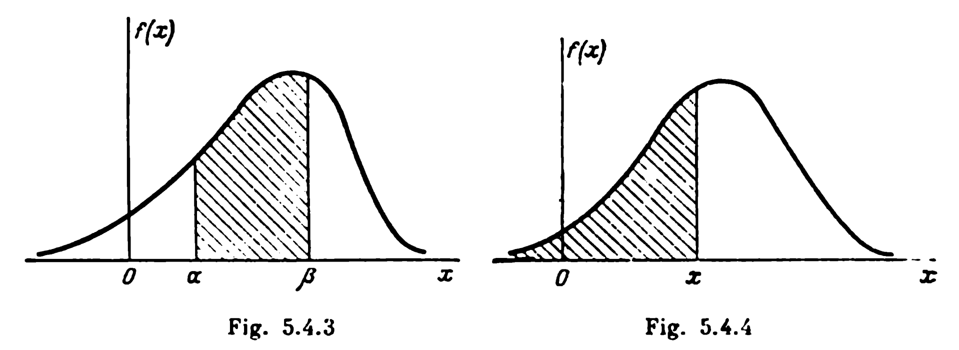


Fig. 5.4.3 Fig. 5.4.4

Enonçons les propriétés essentielles de la densité de probabilité.

1. La densité de probabilité est une fonction non négative:

$$f(x) \geqslant 0.$$

Cette propriété découle immédiatement du fait que la fonction de répartition $F(x)$ est une fonction non décroissante.

2. L'intégrale prise de $-\infty$ à $+\infty$ de la densité de probabilité est égale à l'unité:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

Ceci découle de la formule (5.4.4) et du fait que $F(+\infty) = 1$.

Pour les propriétés essentielles de la densité de probabilité on peut donner l'interprétation géométrique suivante:

1) toute la courbe de densité de probabilité se trouve au-dessus de l'axe des abscisses;

2) l'aire totale comprise entre la courbe de densité de probabilité et l'axe des abscisses est égale à l'unité.

Intéressons-nous aux dimensions de la fonction de répartition et de la densité de probabilité. La fonction de répartition $F(x)$, comme toute probabilité, est une grandeur sans dimension. Quant

à la densité de probabilité $f(x)$, on peut voir à partir de la formule (5.4.1), qu'elle a la dimension inverse de la variable aléatoire.

Exemple 1. La fonction de répartition de la variable aléatoire X est donnée par la formule suivante:

$$F(x) = \begin{cases} 0 & \text{pour } x \leqslant 0; \\ ax^2 & \text{pour } 0 < x \leqslant 1; \\ 1 & \text{pour } x > 1. \end{cases}$$

a) Trouver le coefficient a.
b) Trouver la densité de probabilité $f(x)$.
c) Trouver la probabilité pour la variable X de se trouver dans l'intervalle compris entre 0,25 et 0,5.

Solution. a) La fonction de répartition de X étant continue, pour $x = 1$ on a $ax^2 = 1$, d'où $a = 1$.

b) La densité de probabilité de X est donnée par la formule:

$$f(x) = \begin{cases} 0 & \text{pour } x \leqslant 0; \\ 2x & \text{pour } 0 < x \leqslant 1; \\ 0 & \text{pour } x > 1. \end{cases}$$

c) En vertu de la formule (5.3.1) on a:

$$P(0{,}25 < X < 0{,}5) = F(0{,}5) - F(0{,}25) = 0{,}5^2 - 0{,}25^2 = 0{,}1875.$$

Exemple 2. La variable aléatoire X suit une loi de répartition de densité

$$f(x) = a\cos x \qquad \text{pour } -\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2};$$

$$f(x) = 0 \qquad \text{pour } x < -\frac{\pi}{2} \text{ ou } x > \frac{\pi}{2}.$$

a) Trouver le coefficient a.
b) Tracer le graphique de la densité de la probabilité $f(x)$.
c) Trouver la fonction de répartition $F(x)$ et tracer son graphique.
d) Trouver la probabilité pour la variable X de se trouver entre 0 et $\frac{\pi}{4}$.

Solution. a) Pour trouver le coefficient a nous utilisons la propriété suivante de la fonction de répartition:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a\cos x\,dx = 2a = 1,$$

d'où $a = 1/2$.

b) Le graphique de la densité de probabilité $f(x)$ est donné sur la figure 5.4.5.

c) La formule (5.4.4) permet de trouver l'expression de la fonction de répartition:

$$F(x) = \begin{cases} 0 & \text{pour } x < -\frac{\pi}{2}; \\ 1/2(\sin x + 1) & \text{pour } -\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2}; \\ 1 & \text{pour } x > \frac{\pi}{2}. \end{cases}$$

Sur la figure 5.4.6 on peut voir le graphique de la fonction $F(x)$.

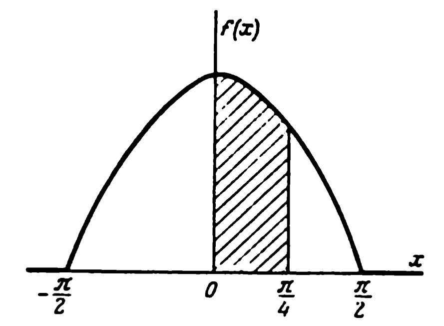


Fig. 5.4.5

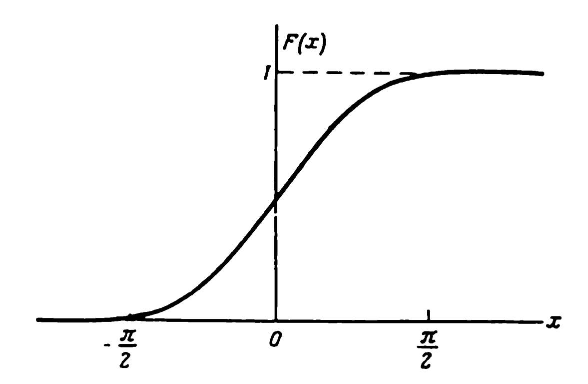


Fig. 5.4.6

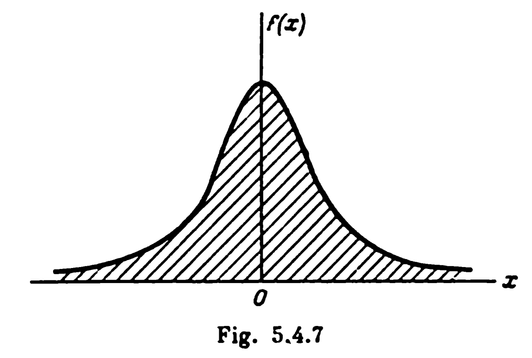


Fig. 5.4.7

d) La formule (5.3.1) donne:

$$P\left(0<X<\frac{\pi}{4}\right)=\frac{1}{2}\left(\sin\frac{\pi}{4}+1\right)-\frac{1}{2}(\sin 0+1)=\frac{\sqrt{2}}{4}.$$

La formule (5.4.3) permet d'obtenir le même résultat mais d'une manière un peu plus compliquée.

Exemple 3. La densité de probabilité de la variable aléatoire X est donnée par la formule suivante:

$$f(x)=\frac{1}{\pi(1+x^2)}\ ^{*)}.$$

a) Tracer le graphique de la densité $f(x)$.

b) Trouver la probabilité pour la variable aléatoire X de se trouver dans l'intervalle $(-1, +1)$.

Solution. a) Le graphique de la densité de probabilité est donné sur la figure 5.4.7.

b) En vertu de la formule (5.4.3) on a:

$$P(-1<X<1)=\int_{-1}^{1}\frac{dx}{\pi(1+x^2)}=\frac{1}{\pi}\operatorname{arc\,tg} x\Big|_{-1}^{1}=\frac{1}{2}.$$

5.5. Caractéristiques numériques des variables aléatoires. Leur rôle et leur désignation

Dans ce chapitre nous avons étudié un certain nombre de caractéristiques des variables aléatoires appelées lois de répartition. Rappelons ces caractéristiques:

pour une variable aléatoire discrète:

a) la fonction de répartition,

b) le tableau de répartition (graphiquement le polygone de répartition);

pour une variable aléatoire continue:

a) la fonction de répartition,

b) la densité de probabilité.

Chaque loi de répartition est une certaine fonction et il suffit d'indiquer cette fonction pour donner une description complète de la variable aléatoire du point de vue probabiliste.

Cependant dans de nombreuses applications pratiques on n'a pas besoin d'une caractéristique aussi complète qu'est la loi de répartition. Souvent il suffit d'indiquer seulement certains paramètres numériques caractérisant dans une certaine mesure les traits essentiels de la répartition étudiée: par exemple, une valeur moyenne autour de laquelle se trouvent groupées les valeurs éventuelles de la variable aléatoire; un nombre quelconque caractérisant la dispersion de ces valeurs autour de la valeur moyenne, etc. En utilisant ces caractéristiques qu'on appelle *numériques*, on peut exprimer

*) Loi dite de Cauchy.

toutes les connaissances dont on dispose sur la variable aléatoire étudiée d'une manière plus compacte, à l'aide d'un nombre minimal de paramètres numériques.

Les caractéristiques numériques et les manipulations qu'on peut effectuer sur elles jouent un rôle de premier plan en théorie des probabilités. L'utilisation des caractéristiques numériques facilite la résolution de nombreux problèmes. Très souvent on arrive à résoudre le problème définitivement sans faire recours aux lois de répartition, se bornant aux seules caractéristiques numériques. Ce qui est important c'est que lorsque dans un problème on a affaire à un grand nombre de variables aléatoires, chacune intervenant d'une certaine manière dans le résultat numérique de l'expérience, la loi de répartition de ce résultat peut en une grande mesure être considérée comme indépendante des lois des différentes variables du problème (on voit ainsi apparaître la loi dite normale). Dans ces cas pour trouver, et ceci avec assez d'exactitude, la loi de répartition de la somme on n'a pas besoin des lois de répartition des variables aléatoires figurant dans le problème ; il suffit d'en connaître certaines caractéristiques numériques.

En théorie des probabilités et en statistique théorique on fait usage d'un grand nombre de caractéristiques numériques qui diffèrent par leur rôle et les domaines d'application. Nous allons étudier ici quelques-unes d'entre elles, utilisées le plus souvent

5.6. Caractéristiques de position (espérance mathématique, mode, médiane)

Parmi les caractéristiques numériques des variables aléatoires il faut tout d'abord distinguer celles qui caractérisent la position de la variable aléatoire sur l'axe numérique, c'est-à-dire celles qui indiquent une certaine valeur moyenne, approchée, autour de laquelle se trouvent groupées toutes les valeurs possibles de la variable aléatoire étudiée.

La valeur moyenne d'une variable aléatoire est un certain nombre en quelque sorte représentatif de cette variable et pouvant la remplacer dans les calculs approchés. Lorsque l'on dit que le temps moyen de fonctionnement d'un tube électronique est égal à 100 heures ou que le point moyen d'impact se trouve déplacé par rapport à la cible de deux mètres à droite, on mentionne une certaine caractéristique numérique de la variable aléatoire, décrivant sa position sur l'axe numérique, c'est-à-dire sa caractéristique de position.

Parmi les caractéristiques de position l'*espérance mathématique* d'une variable aléatoire, appelée parfois *valeur moyenne*, occupe en théorie des probabilités un rôle de choix.

Soit la variable aléatoire discrète X qui prend les valeurs possibles $x_1, x_2, \ldots, x_n$ avec les probabilités respectives $p_1, p_2, \ldots, p_n$.

Il y a lieu de caractériser par un nombre quelconque la position des valeurs de cette variable aléatoire sur l'axe des abscisses, compte tenu du fait que ces valeurs ont des probabilités différentes. A cet effet il est tout naturel d'utiliser les moyennes pondérées des valeurs x_i, les poids étant pris proportionnels aux probabilités de ces valeurs. Ainsi nous calculerons la valeur moyenne de la variable aléatoire X que nous désignerons par $M[X]$:

$$M[X] = \frac{x_1p_1 + x_2p_2 + \ldots + x_np_n}{p_1 + p_2 + \ldots + p_n} = \frac{\sum_{i=1}^{n} x_ip_i}{\sum_{i=1}^{n} p_i},$$

ou compte tenu de $\sum_{i=1}^{n} p_i = 1$,

$$M[X] = \sum_{i=1}^{n} x_ip_i. \quad (5.6.1)$$

Cette moyenne pondérée est appelée espérance mathématique de la variable aléatoire. Nous avons donc introduit une des notions fondamentales de la théorie des probabilités, à savoir la notion d'espérance mathématique.

On appelle espérance mathématique d'une variable aléatoire la somme des produits de toutes les valeurs possibles de cette variable par les probabilités de ces valeurs.

Notons que dans la formulation ci-dessus la définition de l'espérance mathématique n'est en toute rigueur applicable qu aux variables aléatoires discrètes; plus loin nous montrerons comment cette notion se généralise au cas continu.

Pour rendre la notion d'espérance mathématique plus claire nous allons revenir à l'interprétation mécanique de la répartition des valeurs d'une variable aléatoire discrète. Imaginons sur l'axe des abscisses les points d'abscisses $x_1, x_2, \ldots, x_n$ où se trouvent concentrées les masses $p_1, p_2, \ldots, p_n$, avec $\sum_{i=1}^{n} p_i = 1$. Il est alors évident que l'espérance mathématique donnée par la formule (5.6.1) n'est rien d'autre que l'*abscisse du centre de gravité* du système donné de points matériels.

L'espérance mathématique d'une variable aléatoire X est liée à la *moyenne arithmétique* des valeurs observées dans un grand nombre d'expériences par une relation pareille à celle qui existe entre la fréquence relative et la probabilité; plus exactement, pour un grand nombre d'expériences la moyenne arithmétique des valeurs observées d'une variable aléatoire tend (converge en probabilité) vers son espérance mathématique. La relation existant entre la fréquence

et la probabilité entraîne une relation analogue entre la moyenne arithmétique et l'espérance mathématique.

En effet, considérons une variable aléatoire discrète X donnée par le tableau de répartition:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

où $p_i = P(X = x_i)$.

On effectue N expériences indépendantes dans chacune desquelles X prend une certaine valeur. Supposons que la valeur x_1 est observée m_1 fois, la valeur x_2, m_2 fois et en général la valeur x_i, m_i fois. Il est évident que

$$\sum_{i=1}^{n} m_i = N.$$

Calculons la moyenne arithmétique des valeurs observées laquelle à la différence de l'espérance mathématique $M[X]$ sera désignée par $M^*[X]$:

$$M^*[X] = \frac{x_1 m_1 + x_2 m_2 + \ldots + x_n m_n}{N} =$$

$$= x_1 \frac{m_1}{N} + x_2 \frac{m^2}{N} + \ldots + x_n \frac{m_n}{N} = \sum_{i=1}^{n} x_i \frac{m_i}{N}.$$

Mais $\frac{m_i}{N}$ n'est rien d'autre que la fréquence (ou la probabilité statistique) de l'événement $X = x_i$; nous désignerons cette fréquence par p_i^*. On a alors:

$$M^*[X] = \sum_{i=1}^{n} x_i p_i^*,$$

c'est-à-dire que la moyenne arithmétique des valeurs observées d'une variable aléatoire est égale à la somme des produits de toutes les valeurs possibles de cette variable aléatoire par la fréquence de ces valeurs.

Lorsque le nombre d'expériences N augmente les fréquences p_i^* tendent (convergent en probabilité) vers les probabilités correspondantes p_i, par conséquent la moyenne arithmétique $M^*[X]$ des valeurs observées de X tendra (convergera en probabilité) vers son espérance mathématique $M[X]$.

La relation démontrée ci-dessus entre la moyenne arithmétique et l'espérance mathématique est l'une des variétés de la loi des grands nombres. Nous en donnerons la démonstration rigoureuse au chapitre 13.

Nous savons déjà que toutes les formes de la loi des grands nombres sont la constatation du fait de la stabilité de certaines moyennes pour un grand nombre d'expériences. Il s'agit ici de la stabilité de la moyenne arithmétique d'une série d'observations d'une même grandeur. Lorsque le nombre d'expériences n'est pas important, la moyenne arithmétique de leurs résultats est aléatoire; avec augmentation du nombre d'expériences elle a tendance à se stabiliser, en convergeant vers une grandeur constante qu'est l'espérance mathématique.

La stabilité des moyennes pour un grand nombre d'expériences peut facilement être vérifiée expérimentalement. Par exemple, pour trouver le poids exact d'un corps quelconque on le pèse dans un laboratoire sur une balance de précision, chaque pesée fournit une nouvelle valeur; pour diminuer l'erreur d'observation on pèse le corps plusieurs fois et l'on prend la moyenne arithmétique des valeurs relevées. Il est facile de voir que si l'on continue à augmenter le nombre d'expériences (pesées), la moyenne arithmétique est de moins en moins influencée par cette augmentation et lorsque le nombre d'expériences est suffisamment important, elle ne change pratiquement plus.

La formule (5.6.1) de l'espérance mathématique correspond aux cas d'une variable aléatoire discrète. Pour une variable continue X l'espérance mathématique n'est évidemment plus une somme, mais une intégrale, soit:

$$M[X] = \int_{-\infty}^{\infty} x f(x)\, dx, \qquad (5.6.2)$$

où $f(x)$ est la densité de probabilité de X.

La formule (5.6.2) peut être obtenue à partir de la formule (5.6.1) à condition d'y remplacer les valeurs x_i par le paramètre x variant d'une manière continue, les probabilités correspondantes p_i par l'élément de probabilité $f(x)\,dx$ et la somme finie par une intégrale. Nous utiliserons souvent dans la suite cette méthode de généralisation des formules démontrées pour le cas des variables discrètes.

Dans l'interprétation mécanique, l'espérance mathématique d'une variable aléatoire continue conserve le même sens, celui de l'abscisse du centre de gravité d'une barre dont la masse est répartie d'une manière continue avec la densité $f(x)$. Cette interprétation permet souvent de trouver l'espérance mathématique sans procéder au calcul de l'intégrale (5.6.2), ceci à partir des considérations physiques simples.

Ci-dessus nous avons introduit la désignation $M[X]$ pour l'espérance mathématique de X. Dans certains cas, lorsque la grandeur $M[X]$ entre dans les formules comme un nombre donné, il est plus commode de la désigner par une seule lettre. Nous adoptons la dési-

gnation m_x:

$$m_x = M[X].$$

Nous employerons dans la suite les désignations m_x et $M[X]$ pour l'espérance mathématique parallèlement, en choisissant celle qui se trouve être la plus commode.

Il faut noter que l'espérance mathématique, caractéristique essentielle de position, n'existe pas pour toutes les variables aléatoires. On peut trouver des exemples de variables aléatoires n'ayant pas d'espérance mathématique, car la somme ou l'intégrale correspondantes sont divergentes.

Considérons par exemple la variable aléatoire discrète X donnée par le tableau de répartition suivant:

x_i	2	2^2	...	2^i	...
p_i	$\frac{1}{2}$	$\frac{1}{2^2}$	...	$\frac{1}{2^i}$	...

Il est facile de vérifier que $\sum_{i=1}^{n} p_i = 1$, cependant la somme $\sum_{i=1}^{\infty} x_i p_i$ est ici divergente et par conséquent l'espérance mathématique de X n'existe pas. De tels cas n'ont pas d'intérêt pratique car généralement les variables aléatoires auxquelles on a affaire ont des valeurs limitées et leur espérance mathématique existe sans aucun doute.

Ci-dessus nous avons donné les formules pour l'espérance mathématique des variables aléatoires discrètes (5.6.1) et continues (5.6.2).

Si la variable X est du type mixte, son espérance mathématique est donnée par la formule:

$$M[X] = \sum_i x_i p_i + \int xF'(x)\,dx, \qquad (5.6.3)$$

la somme s'étendant à tous les points x_i où la fonction de répartition admet une discontinuité, et l'intégrale, à tous le intervalles de répartition continue.

En plus de l'espérance mathématique, caractéristique essentielle de position d'une variable aléatoire, on utilise parfois d'autres caractéristiques de position, en particulier le *mode* et la *médiane*.

On appelle *mode* d'une variable aléatoire sa valeur la plus probable. Le terme de « valeur la plus probable » n'est applicable en toute rigueur qu'aux variables discrètes; pour une variable continue le mode est la valeur pour laquelle la densité de probabilité est maximale. On désigne le mode par la lettre $\mathcal{M}$. Sur la figure 5.6.1 on peut voir le mode dans le cas discret et sur la figure 5.6.2 dans le cas continu.

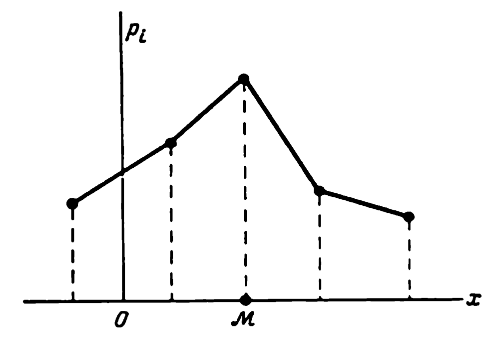


Fig. 5.6.1

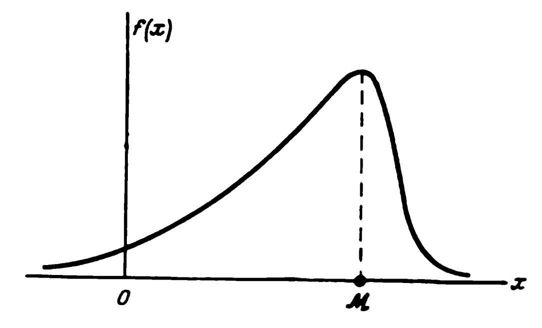


Fig. 5.6.2

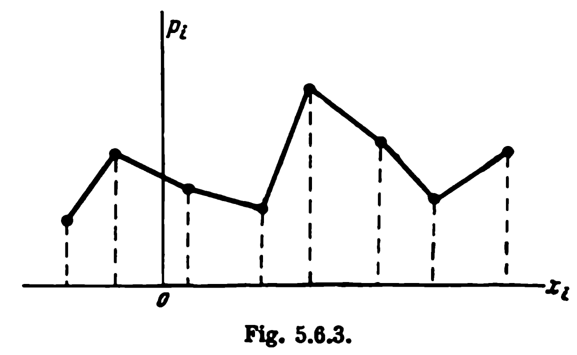


Fig. 5.6.3.

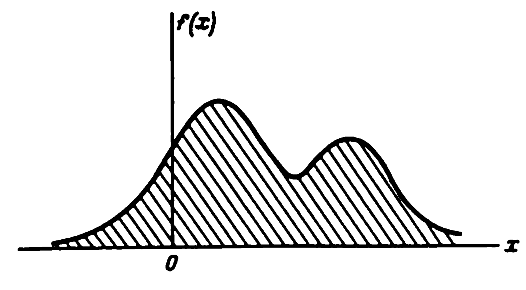


Fig. 5.6.4

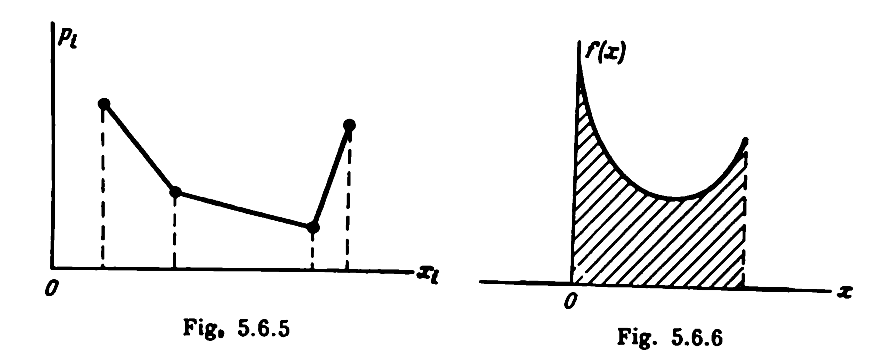


Fig. 5.6.5

Fig. 5.6.6

Fig. 5.6.7

Si le polygone de la répartition (densité de probabilité) a plus d'un maximum, la répartition est dite *polymodale* (fig. 5.6.3 et 5.6.4).

Parfois on rencontre des répartitions ayant non pas un maximum mais un minimum (fig. 5.6.5 et 5.6.6). Ces répartitions sont dites *antimodales*. A titre d'exemple de répartition antimodale on peut citer la loi obtenue dans l'exemple 5 du § 5.1.

Dans le cas général le mode et l'espérance mathématique d'une variable aléatoire ne coïncident pas. Cependant dans le cas particulier, où la répartition est symétrique et unimodale (c'est-à-dire a un mode) et où l'espérance mathématique existe, celle-ci coïncide avec le mode et le centre de symétrie de la répartition.

Mentionnons encore une caractéristique de position, à savoir la *médiane* d'une variable aléatoire. Il est d'usage d'employer cette caractéristique seulement pour des variables aléatoires continues, bien que du point de vue formel elle puisse également être utilisée pour une variable discrète.

On appelle *médiane* d'une variable aléatoire X sa valeur $\mathcal{M}e$ telle que

$$P(X < \mathcal{M}e) = P(X > \mathcal{M}e),$$

c'est-à-dire que la variable aléatoire considérée a même probabilité d'être inférieure ou supérieure à $\mathcal{M}e$. Du point de vue géométrique, la médiane est l'abscisse du point divisant en deux parties égales l'aire limitée par la courbe $f(x)$ et l'axe des abscisses (fig. 5.6.7).

Dans le cas d'une répartition modale et symétrique, la médiane coïncide avec l'espérance mathématique et le mode.

5.7. Moments. Variance. Ecart quadratique moyen

En plus des moyennes, qui sont les caractéristiques de position, on utilise également certaines autres caractéristiques des répartitions. Les *moments* sont les plus courantes de ces caractéristiques.

En mécanique on utilise couramment la notion de moment pour décrire la répartition des masses (moments statistiques, moments d'inertie, etc.). En théorie des probabilités la notion de moment a à peu près la même signification. Dans les applications pratiques le plus souvent ou utilise les moments initiaux et les moments centrés.

On appelle *moment initial d'ordre s* d'une variable aléatoire discrète X et on note $\alpha_s[X]$ la somme

$$\alpha_s[X] = \sum_{i=1}^{n} x_i^s p_i. \tag{5.7.1}$$

Il est facile de voir que cette définition coïncide avec celle du moment initial d'ordre s utilisée en mécanique si l'on considère que les

masses $p_1, p_2, \ldots, p_n$ se trouvent concentrées aux points $x_1, x_2, \ldots, x_n$.

Dans le cas d'une variable aléatoire continue X on appelle moment initial d'ordre s l'intégrale

$$\alpha_s[X] = \int_{-\infty}^{\infty} x^s f(x)\,dx. \tag{5.7.2}$$

On voit aisément que la caractéristique principale de position introduite dans le paragraphe précédent, à savoir l'espérance mathématique, n'est rien d'autre que le *moment initial d'ordre* 1 de X.

En utilisant le signe d'espérance mathématique on peut réunir en une seule les formules (5.7.1) et (5.7.2). En effet, les formules (5.7.1) et (5.7.2) ont la même structure que (5.6.1) et (5.6.2), avec cette seule différence qu'à la place de x_i et x on a respectivement x_i^s et x^s. On peut donner alors une définition générale du moment d'ordre s, applicable tant pour les variables discrètes que continues: *on appelle moment initial d'ordre s de X l'espérance mathématique de la s-ième puissance de X*, soit *):

$$\alpha_s[X] = M[X^s]. \tag{5.7.3}$$

Avant de donner la définition du moment centré introduisons la notion de variable aléatoire centrée.

Soit une variable aléatoire X d'espérance mathématique m_x. On appelle *variable aléatoire centrée* associée à X et on note $\overset{\circ}{X}$ la différence

$$\overset{\circ}{X} = X - m_x. \tag{5.7.4}$$

Il est facile de voir que l'espérance mathématique d'une variable aléatoire centrée est nulle. En effet pour une variable discrète on a:

$$M[\overset{\circ}{X}] = M[X - m_x] = \sum_{i=1}^{n} (x_i - m_x)\,p_i =$$

$$= \sum_{i=1}^{n} x_i p_i - m_x \sum_{i=1}^{n} p_i = m_x - m_x = 0. \tag{5.7.5}$$

Dans le cas continu on a le même résultat.

Le centrage d'une variable aléatoire est de toute évidence équivalent à la translation de l'origine au point moyen, « central », dont l'abscisse est égale à l'espérance mathématique.

Les moments d'une variable aléatoire centrée sont appelés *moments centrés*. Ils sont analogues aux moments par rapport au centre de gravité en mécanique.

---

*) Nous préciserons dans le chapitre 10 la notion d'espérance mathématique d'une fonction d'une variable aléatoire.

Ainsi, on appelle *moment centré d'ordre s* de X l'espérance mathématique de la s-ième puissance de la variable centrée associée à X:

$$\mu_s[X] = M[\overset{\circ}{X}{}^s] = M[X - m_x)^s]. \quad (5.7.6)$$

Pour une variable aléatoire discrète le moment centré d'ordre s est donné par la somme

$$\mu_s = \sum_{i=1}^{n} (x_i - m_x)^s p_i, \quad (5.7.7)$$

et pour une variable aléatoire continue par l'intégrale

$$\mu_s = \int_{-\infty}^{\infty} (x - m_x)^s f(x)\, dx. \quad (5.7.8)$$

Dans la suite, quand il ne peut y avoir de doute quant à la grandeur à laquelle correspond le moment en question, pour simplifier l'écriture on écrira simplement α_s et μ_s au lieu de $\alpha_s[X]$ et $\mu_s[X]$.

Il est évident que pour toute variable aléatoire le *moment centré d'ordre 1 est nul*:

$$\mu_1 = M[\overset{\circ}{X}] = M[X - m_x] = 0, \quad (5.7.9)$$

car l'espérance mathématique d'une variable centrée est toujours nulle.

Cherchons les relations existant entre les moments centrés et initiaux de différents ordres. La démonstration sera donnée seulement pour le cas des variables discrètes; il est facile de voir que des relations identiques existent pour les variables continues, pour s'en convaincre il suffit de remplacer les sommes finies par des intégrales et les probabilités par des éléments de probabilité.

Considérons maintenant le moment centré deux:

$$\mu_2 = M[\overset{\circ}{X}{}^2] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i =$$
$$= \sum_{i=1}^{n} x_i^2 p_i - 2m_x \sum_{i=1}^{n} x_i p_i + m_x^2 \sum_{i=1}^{n} p_i = \alpha_2 - 2m_x^2 + m_x^2 = \alpha_2 - m_x^2.$$

D'une manière analogue pour le moment centré trois on obtient:

$$\mu_3 = M[\overset{\circ}{X}{}^3] = \sum_{i=1}^{n} (x_i - m_x)^3 p_i = \sum_{i=1}^{n} x_i^3 p_i - 3m_x \sum_{i=1}^{n} x_i^2 p_i +$$
$$+ 3m_x^2 \sum_{i=1}^{n} x_i p_i - m_x^3 \sum_{i=1}^{n} p_i = \alpha_3 - 3\alpha_2 m_x + 2m_x^3.$$

Les expressions pour μ_4, μ_5, etc. peuvent être obtenues d'une manière analogue.

Pour les moments centrés d'une variable aléatoire quelconque X on a donc:

$$\left.\begin{array}{l} \mu_1 = 0, \\ \mu_2 = \alpha_2 - m_x^2, \\ \mu_3 = \alpha_3 - 3m_x\alpha_2 + 2m_x^3, \\ \dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots \end{array}\right\} \qquad (5.7.10)$$

Dans le cas général on peut envisager les moments non seulement par rapport à l'origine des coordonnées (moments initiaux) ou à l'espérance mathématique (moments centrés), mais également par rapport à un point quelconque a:

$$\gamma_s = M[(X-a)^s]. \qquad (5.7.11)$$

Cependant les moments centrés ont devant tous les autres l'avantage suivant: le premier moment centré est toujours nul et le moment centré deux est minimal. Nous allons démontrer cette dernière assertion. Pour une variable aléatoire discrète X, pour $s = 2$, la formule (5.7.11) s'écrit:

$$\gamma_2 = \sum_{i=1}^{n} (x_i - a)^2 p_i. \qquad (5.7.12)$$

On peut transformer cette expression de la manière suivante:

$$\gamma_2 = \sum_{i=1}^{n} (x_i - m_x + m_x - a)^2 p_i = \sum_{i=1}^{n} (x_i - m_x)^2 p_i -$$

$$- 2(m_x - a) \sum_{i=1}^{n} (x_i - m_x) p_i + (m_x - a)^2 = \mu_2 + (m_x - a)^2.$$

Il est évident que cette grandeur est minimale pour $m_x = a$, c'est-à-dire lorsque le moment est pris par rapport au point m_x.

Pour caractériser une variable aléatoire on utilise le plus souvent le moment initial d'ordre 1 (espérance mathématique) $m_x = \alpha_1$ et le moment centré deux μ_2.

Le moment centré deux de X s'appelle *variance* de X. Vu l'importance de cette caractéristique on introduit pour elle la désignation spéciale $D[X]$:

$$\mu_2 = D[X].$$

Par définition du moment centré, on a:

$$D[X] = M[\overset{\circ}{X}{}^2], \qquad (5.7.13)$$

c'est-à-dire que la *variance de X est l'espérance mathématique du carré de la variable centrée associée.*

En remplaçant dans l'expression (5.7.13) la grandeur $\overset{\circ}{X}$ par son expression on a:

$$D[X] = M[(X - m_x)^2]. \tag{5.7.14}$$

Pour le calcul direct de la variance on peut utiliser les formules:

$$D[X] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i, \tag{5.7.15}$$

$$D[X] = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)\, dx \tag{5.7.16}$$

pour les variables discrètes et continues respectivement.

La variance d'une variable aléatoire caractérise la *dispersion* de ses valeurs au voisinage de son espérance mathématique.

En faisant appel à l'interprétation mécanique de la répartition on voit que la variance n'est rien d'autre que le *moment d'inertie* de la répartition en question des masses par rapport au centre de gravité (à l'espérance mathématique).

La variance a la dimension du carré de la variable aléatoire. Pour mieux caractériser la dispersion, plus commode est d'utiliser une grandeur dont la dimension est celle de la variable aléatoire. A cet effet, on prend la racine carrée de la variance. La grandeur ainsi obtenue est appelée *écart quadratique moyen* de X ou *écart type* de X. On désigne l'écart quadratique moyen par $\sigma[X]$:

$$\sigma[X] = \sqrt{D[X]}. \tag{5.7.17}$$

Pour simplifier l'écriture on utilise souvent des désignations abrégées σ_x et D_x de l'écart quadratique moyen et de la variance. Quand il ne peut y avoir confusion quant aux grandeurs auxquelles ces caractéristiques se rapportent, on omet l'indice x, en écrivant simplement σ et D.

Dans les applications pratiques on exprime souvent la variance d'une variable aléatoire en fonction de son moment initial deux [seconde formule (5.7.10)]:

$$D_x = \alpha_2 - m_x^2. \tag{5.7.18}$$

L'espérance mathématique m_x et la variance D_x (ou l'écart quadratique moyen σ_x) sont les caractéristiques le plus souvent utilisées d'une répartition. Elles caractérisent ses traits essentiels, à savoir sa position et son degré de dispersion. Pour en donner une description plus détaillée, on fait appel à des moments d'ordre supérieur.

Le moment centré trois caractérise l'asymétrie d'une répartition. Si la répartition est symétrique par rapport à l'espérance mathématique (ou, dans l'interprétation mécanique, la masse est répartie

symétriquement par rapport au centre de gravité), tous les moments impairs (s'ils existent) sont nuls. En effet, dans la somme

$$\mu_s = \sum_{i=1}^{n} (x_i - m_x)^s p_i,$$

lorsque la loi de répartition est symétrique par rapport à m_x et s est impair, à chaque terme positif correspond un terme négatif égal en valeur absolue, de sorte que la somme est nulle. Il en est de même pour l'intégrale

$$\mu_s = \int_{-\infty}^{\infty} (x - m_x) \; f(x)\, dx,$$

qui est nulle, car c'est une intégrale d'une fonction impaire prise dans des limites symétriques.

Il est donc naturel pour caractériser l'asymétrie d'une répartition de prendre l'un des moments impairs, dont le plus simple est le

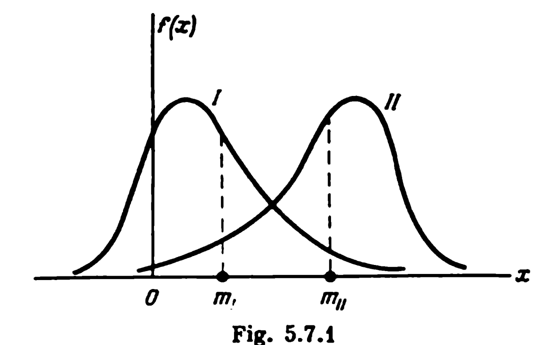


Fig. 5.7.1

moment d'ordre trois. Il a la dimension du cube de la variable aléatoire; pour obtenir une caractéristique sans dimension on divise le moment trois μ_3 par le cube de l'écart quadratique moyen. La grandeur ainsi obtenue est appelée *coefficient d'asymétrie* ou simplement asymétrie; nous désignerons cette grandeur par Sk:

$$Sk = \frac{\mu_3}{\sigma^3}. \tag{5.7.19}$$

Sur la figure 5.7.1 on peut voir deux répartitions asymétriques, l'une d'elles (courbe *I*) a une asymétrie positive ($Sk > 0$); l'autre (courbe *II*) est à asymétrie négative ($Sk < 0$).

Le moment centré quatre sert à caractériser l'aplatissement de la courbe de répartition, c'est-à-dire montre combien la répartition est pointue ou aplatie. Cette propriété de la répartition est donnée par le *coefficient d'aplatissement*. On appelle coefficient d'aplatisse-

ment d'une répartition et on note Ex la grandeur

$$Ex = \frac{\mu_4}{\sigma^4} - 3. \qquad (5.7.20)$$

On retranche le nombre 3 du rapport $\frac{\mu_4}{\sigma^4}$, car pour le cas très important que l'on rencontre dans les applications pratiques, cas de la loi normale (que nous étudierons en détail plus loin), on a $\frac{\mu_4}{\sigma^4} =$ $= 3$. Donc le coefficient d'aplatissement de la loi normale est égal

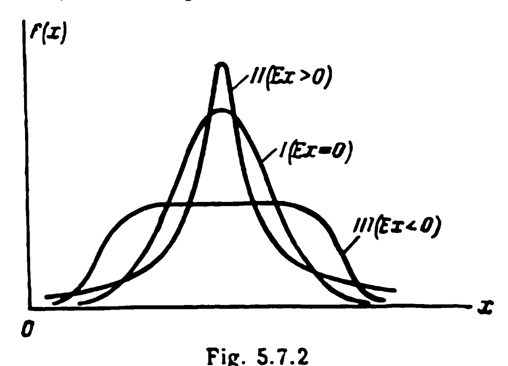


Fig. 5.7.2

à zero; les courbes plus pointues ont un coefficient d'aplatissement positif; les courbes plus aplaties ont un coefficient d'aplatissement négatif.

Sur la figure 5.7.2 on peut voir une répartition normale (courbe *I*), une répartition à coefficient d'aplatissement positif (courbe *II*) et une répartition à coefficient d'aplatissement négatif (courbe *III*).

En plus des moments initiaux et centrés, on utilise parfois dans les applications pratiques les *moments absolus* correspondants, donnés par les expressions:

$$\beta_s = M[|X|^s]; \quad \nu_s = M[|\overset{\circ}{X}|^s].$$

On remarque que les moments absolus d'ordres pairs coïncident avec les moments ordinaires.

Parmi les moments absolus le plus employé est le moment absolu centré d'ordre 1:

$$\nu_1 = M[|\overset{\circ}{X}|] = M[|X - m_x|], \qquad (5.7.21)$$

appelé *écart arithmétique moyen*. Tout comme la variance et l'écart quadratique moyen, l'écart arithmétique moyen caractérise également la dispersion d'une répartition.

L'espérance mathématique, le mode, la médiane, les moments initiaux et centrés et, en particulier, la variance, l'écart quadratique

moyen, les coefficients d'asymétrie et d'aplatissement sont les caractéristiques numériques les plus usuelles des variables aléatoires. Dans de nombreux problèmes il arrive qu'on ne peut pas déterminer la loi de répartition ou bien on n'en a pas besoin. On se limite alors à la description approchée de la variable aléatoire à l'aide des caractéristiques numériques, chacune d'elles exprimant une des propriétés essentielles de la répartition.

On se sert souvent des caractéristiques numériques en approchant une répartition par une autre; on s'efforce, en général, de conserver certains moments importants.

Exemple 1. On effectue une expérience dans laquelle l'événement A peut être réalisé avec la probabilité p, ou ne pas être réalisé. Soit la variable aléatoire X représentant le nombre de réalisations de l'événement A (variable indicatrice de l'événement A). Trouver son espérance mathématique, sa variance et son écart quadratique moyen.

Solution. Le tableau de répartition de X est:

x_i	0	1
p_i	q	p

où $q = 1 - p$ est la probabilité de la non-réalisation de l'événement A.

L'espérance mathématique de X est donnée par la formule (5.6.1):

$$m_x = M[X] = 0\cdot q + 1\cdot p = p,$$

et la variance de X par la formule (5.7.15):

$$D_x = D[X] = (0-p)^2\cdot q + (1-p)^2\, p = pq,$$

d'où

$$\sigma_x = \sqrt{pq}.$$

(Le lecteur peut obtenir le même résultat en exprimant la variance en fonction du moment initial deux).

Exemple 2. Trois coups indépendants sont tirés sur une cible; la probabilité d'atteindre le but avec chaque coup est égale à 0,4. Soit la variable aléatoire X représentant le nombre de coups réussis. Trouver l'espérance mathématique, la variance, l'écart quadratique moyen et le coefficient d'asymétrie de X.

Solution. Le tableau de répartition de X est:

x_i	0	1	2	3
p_i	0,216	0,432	0,288	0,064

Le calcul des caractéristiques numériques de X donne:

$$m_x = 0\cdot 0{,}216 + 1\cdot 0{,}432 + 2\cdot 0{,}288 + 3\cdot 0{,}064 = 1{,}2;$$

$$D_x = (0 - 1{,}2)^2\cdot 0{,}216 + (1 - 1{,}2)^2\cdot 0{,}432 + (2 - 1{,}2)^2\cdot 0{,}288 + (3 - 1{,}2)^2\cdot 0{,}064 = 0{,}72;$$

$$\sigma_x = \sqrt{D_x} = \sqrt{0{,}72} = 0{,}848;$$

$$\mu_3 = (0 - 1{,}2)^3\cdot 0{,}216 + (1 - 1{,}2)^3\cdot 0{,}432 + (2 - 1{,}2)^3\cdot 0{,}288 + (3 - 1{,}2)^3\cdot 0{,}064 = 0{,}144;$$

$$Sk = \frac{\mu_3}{\sigma_x^3} = \frac{0{,}144}{0{,}72\cdot 0{,}848} \approx 0{,}236.$$

Notons qu'on aurait pu calculer ces mêmes caractéristiques bien plus facilement en appliquant les théorèmes sur les caractéristiques numériques des fonctions (voir chapitre 10).

Exemple 3. On effectue une série d'expériences indépendantes jusqu'à la première réalisation de l'événement A (voir exemple 3 du § 5.1). Dans chaque expérience la probabilité de l'événement A est égale à p. Trouver l'espérance mathématique, la variance et l'écart quadratique moyen du nombre X d'expériences à effectuer.

Solution. Le tableau de répartition de X est:

x_i	1	2	3	...	i	...
p_i	p	qp	q^2p	...	$q^{i-1}p$	...

L espérance mathématique de X est donnée par la somme de la série:

$$m_x = 1\cdot p + 2\cdot qp + 3\cdot q^2p + \ldots + i\cdot q^{i-1}p + \ldots =$$
$$= p(1 + 2q + 3q^2 + \ldots + iq^{i-1} + \ldots).$$

Il est facile de voir que l'expression entre parenthèses provient de la dérivation de la progression géométrique suivante:

$$q + q^2 + q^3 + \ldots + q^i + \ldots = \frac{q}{1-q}$$

Par conséquent:

$$1 + 2q + 3q^2 + \ldots + iq^{i-1} + \ldots = \frac{d}{dq}\frac{1}{1-q} = \frac{1}{(1-q)^2} = \frac{1}{p^2},$$

d'où

$$m_x = \frac{p}{p^2} = \frac{1}{p}$$

Pour déterminer la variance de X, calculons d'abord son moment initial deux:

$$\alpha_2 = \sum_{i=1}^{\infty} x_i^2 p_i = 1^2\cdot p + 2^2\cdot qp + 3^2\cdot q^2p + \ldots + i^2\cdot q^{i-1}p + \ldots =$$
$$= p(1^2 + 2^2q + 3^2q^2 + \ldots + i^2q^{i-1} + \ldots).$$

Pour calculer la série se trouvant entre parenthèses on multiplie par q la série

$$1 + 2q + 3q^2 + \ldots + iq^{i-1} + \ldots = \frac{1}{(1-q)^2}.$$

On obtient:

$$q + 2q^2 + 3q^3 + \ldots + iq^i + \ldots = \frac{q}{(1-q)^2}.$$

La dérivation de cette série par rapport à q donne:

$$1^2 + 2^2q + 3^2q^2 + \ldots + i^2q^{i-1} + \ldots = \frac{d}{dq}\frac{q}{(1-q)^2} = \frac{1+q}{(1-q)^3}.$$

En multipliant par $p = 1 - q$ on obtient:

$$\alpha_2 = \frac{1+q}{(1-q)^2}.$$

La variance est donnée par la formule (5.7.18):

$$D_x = \alpha_2 - m_x^2 = \frac{1+q}{(1-q)^2} - \frac{1}{(1-q)^2} = \frac{q}{(1-q)^2} = \frac{q}{p^2},$$

d'où

$$\sigma_x = \sqrt{D_x} = \frac{\sqrt{q}}{p}.$$

Exemple 4. La variable aléatoire continue X suit une loi de répartition de densité

$$f(x) = Ae^{-|x|}$$

(fig. 5.7.3).

Trouver le coefficient A. Déterminer l'espérance mathématique, la variance, les coefficients d'asymétrie et d'aplatissement de la répartition de X.

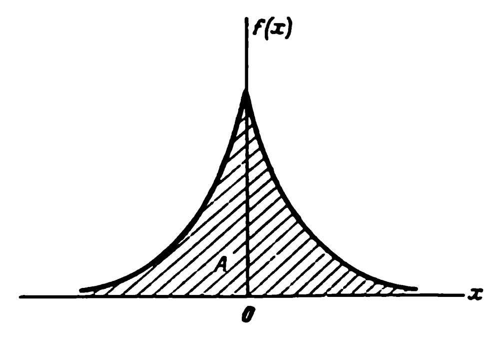


Fig. 5.7.3

Solution. Pour trouver A on utilise la propriété suivante de la densité de probabilité:

$$\int_{-\infty}^{\infty} f(x)\,dx = 2A\int_{0}^{\infty} e^{-x}\,dx = 2A = 1;$$

d'où

$$A = 1/2.$$

La fonction $xe^{-|x|}$ étant impaire, l'espérance mathématique de X est nulle:

$$m_x = \int_{-\infty}^{\infty} \frac{1}{2} xe^{-|x|}\,dx = 0.$$

La variance et l'écart quadratique moyen sont respectivement:

$$D_x = 2\int_{0}^{\infty} \frac{1}{2} x^2 e^{-x}\,dx = 2;$$

$$\sigma_x = \sqrt{D_x} = \sqrt{2}.$$

La répartition étant symétrique on a $Sk = 0$.

Pour calculer le coefficient d'aplatissement il y a lieu de trouver μ_4:

$$\mu_4 = 2\int_{0}^{\infty} \frac{1}{2} x^4 e^{-x}\,dx = 24,$$

d'où

$$Ex = \frac{\mu_4}{\sigma_x^4} - 3 = 3.$$

E x e m p l e 5. La variable aléatoire X est répartie suivant une loi dont la densité est donnée graphiquement sur la figure 5.7.4.

Ecrire l'expression de la densité de probabilité. Trouver l'espérance mathématique, la variance, l'écart quadratique moyen et l'aplatissement de la répartition en question.

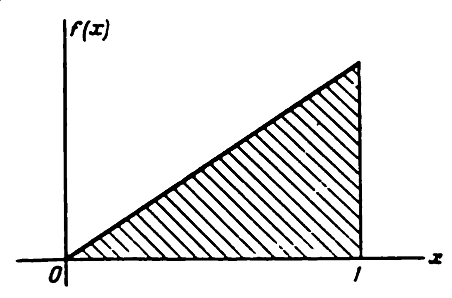


Fig. 5.7.4

S o l u t i o n. L'expression de la densité de probabilité est:

$$f(x) = \begin{cases} ax & \text{pour } 0 < x < 1, \\ 0 & \text{pour } x < 0 \text{ ou } x > 1. \end{cases}$$

En utilisant la propriété de la densité de probabilité d'une répartition on trouve $a = 2$.

L'espérance mathématique de X est:

$$m_x = \int_0^1 2x^2\,dx = \frac{2}{3}.$$

La variance peut s'exprimer en fonction du moment initial deux:

$$\alpha_2 = 2\int_0^1 x^3\,dx = \frac{1}{2}; \qquad D_x = \alpha_2 - m_x^2 = \frac{1}{18},$$

d'où

$$\sigma_x = \frac{1}{3\sqrt{2}}.$$

Le moment initial trois est égal à

$$\alpha_3 = 2\int_0^1 x^4\,dx = \frac{2}{5}.$$

En utilisant la troisième des formules (5.7.10) donnant μ_3 en fonction des moments initiaux, on a:

$$\mu_3 = \alpha_3 - 3m_x\alpha_2 + 2m_x^3 = -\frac{1}{135},$$

d'où

$$Sk = \frac{\mu_3}{\sigma_x^3} = -\frac{2}{5}\sqrt{2}.$$

5.8. Loi de densité uniforme

Dans certains problèmes on rencontre des variables aléatoires continues que l'on sait d avance se trouver dans les limites d'un certain intervalle; de plus on sait que, dans cet intervalle, toutes les valeurs de la variable étudiée sont équiprobables (plus exactement ont même densité de probabilité). On dit que ces variables aléatoires sont réparties suivant la *loi de densité uniforme*.

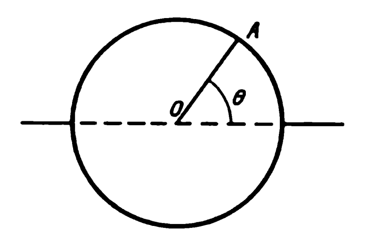


Fig. 5.8.1

Voici quelques exemples de variables aléatoires de densité uniforme.

E x e m p l e 1. On pèse un corps à l'aide d'une balance de précision, mais on ne dispose que de poids multiples d'un gramme; les résultats des pesées montrent que le poids du corps se trouve entre k et $k + 1$ grammes. On adopte que ce poids est égal à $(k + 1/2)$ grammes. L'erreur commise X est de toute évidence une variable aléatoire, répartie avec une densité uniforme sur l'intervalle $(-1/2, 1/2)$ g.

E x e m p l e 2. On fait tourner une roue verticale (fig. 5.8.1), puis cette roue s'arrête par frottement. Considérons la variable aléatoire θ, l'angle que fait après l'arrêt un rayon donné OA avec l'horizontale. Il est évident que la variable θ est répartie avec une densité uniforme sur l'intervalle $(0, 2\pi)$.

E x e m p l e 3. Les rames de métro se suivent avec des intervalles de 2 minutes. Les voyageurs arrivent sur le quai à des instants aléatoires. Le temps T que doit attendre un voyageur est une variable aléatoire, répartie avec une densité uniforme sur l'intervalle (0, 2) minutes.

Considérons une variable aléatoire X suivant une loi de densité uniforme sur l'intervalle (α, β) (fig. 5.8.2) et écrivons l'expression de la densité de probabilité $f(x)$. Cette fonction est constante et égale à c sur l'intervalle (α, β); à l'extérieur de cet intervalle elle est nulle:

$$f(x) = \begin{cases} c & \text{pour} \quad \alpha < x < \beta, \\ 0 & \text{pour} \quad x < \alpha \text{ ou } x > \beta. \end{cases}$$

L'aire comprise entre la courbe $f(x)$ et l'axe des abscisses étant égale à l'unité:

$$c(\beta - \alpha) = 1,$$

on en tire:

$$c = \frac{1}{\beta - \alpha}$$

et la densité de probabilité $f(x)$ est donc:

$$\left.\begin{array}{ll} f(x)=\frac{1}{\beta-\alpha} & \text{pour } \alpha<x<\beta, \\ f(x)=0 & \text{pour } x<\alpha \text{ ou } x>\beta. \end{array}\right\} \qquad (5.8.1)$$

La formule (5.8.1) est celle de la loi de densité uniforme sur l'intervalle (α, β).

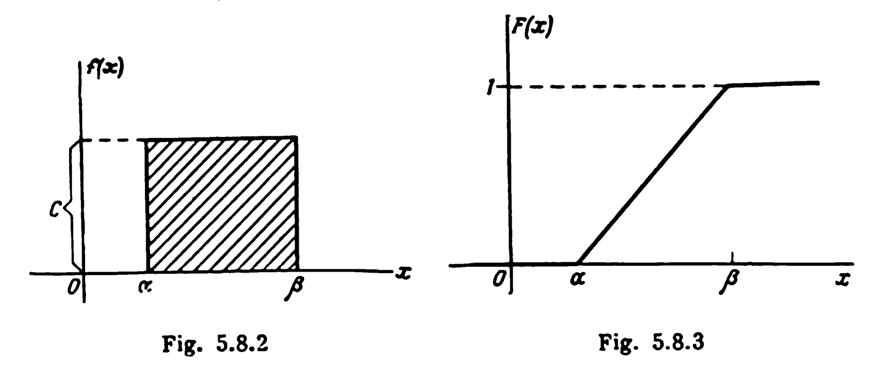


Fig. 5.8.2

Fig. 5.8.3

Ecrivons l'expression de la fonction de répartition $F(x)$. La fonction de répartition est donnée par l'aire comprise entre la courbe $f(x)$ et l'axe des abscisses se trouvant à gauche du point x. Par conséquent:

$$F(x)=\begin{cases} 0 & \text{pour } x<\alpha, \\ \frac{x-\alpha}{\beta-\alpha} & \text{pour } \alpha<x<\beta, \\ 1 & \text{pour } x>\beta. \end{cases}$$

Le graphique de la fonction $F(x)$ est donné sur la figure 5.8.3.

Déterminons certaines des caractéristiques numériques de la variable aléatoire X suivant la loi de répartition uniforme sur l'intervalle (α, β).

L'espérance mathématique de X s'exprime:

$$m_x=\int_{\alpha}^{\beta} \frac{x}{\beta-\alpha}\, dx=\frac{\alpha+\beta}{2}. \qquad (5.8.2)$$

Par suite de la symétrie de la loi uniforme, la médiane de X est aussi égale à $\frac{\alpha+\beta}{2}$.

La loi uniforme n'a pas de mode.

La variance de X est donnée par la formule (5.7.16):

$$D_x = \mu_2 = \frac{1}{\beta-\alpha} \int_\alpha^\beta \left(x - \frac{\alpha+\beta}{2}\right)^2 dx = \frac{(\beta-\alpha)^2}{12}, \tag{5.8.3}$$

d'où l'écart quadratique moyen

$$\sigma_x = \sqrt{D_x} = \frac{\beta-\alpha}{2\sqrt{3}}. \tag{5.8.4}$$

Comme la loi uniforme est symétrique, son coefficient d'asymétrie est nul:

$$Sk = \frac{\mu_3}{\sigma_x^3} = 0. \tag{5.8.5}$$

Pour trouver le coefficient d'aplatissement il y a lieu de calculer préalablement le moment central quatre:

$$\mu_4 = \frac{1}{\beta-\alpha} \int_\alpha^\beta \left(x - \frac{\alpha+\beta}{2}\right)^4 dx = \frac{(\beta-\alpha)^4}{80},$$

d'où

$$Ex = \frac{\mu_4}{\sigma_x^4} - 3 = -1{,}2. \tag{5.8.6}$$

L'écart arithmétique moyen est:

$$\nu_1 = \frac{1}{\beta-\alpha} \int_\alpha^\beta \left| x - \frac{\alpha+\beta}{2} \right| dx = \frac{2}{\beta-\alpha} \int_{\frac{\alpha+\beta}{2}}^\beta \left(x - \frac{\alpha+\beta}{2}\right) dx = \frac{\beta-\alpha}{4}. \tag{5.8.7}$$

Enfin, calculons la probabilité pour la variable aléatoire X répartie suivant la loi uniforme sur l'intervalle (α, β) de tomber dans

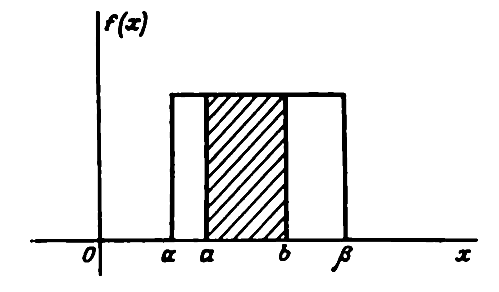


Fig. 5.8.4

l'intervalle (a, b) appartenant à (α, β) (fig. 5.8.4). Du point de vue géométrique cette probabilité est donnée par l'aire hachurée sur

la figure 5.8.4. Il est évident que

$$P(a < X < b) = \frac{b-a}{\beta-\alpha}, \tag{5.8.8}$$

c'est-à-dire que cette probabilité est égale au rapport de la longueur de l'intervalle (a, b) à toute la longueur de l'intervalle (α, β) sur lequel la répartition uniforme est donnée.

5.9. Loi de Poisson

Dans les applications pratiques on rencontre des variables aléatoires réparties suivant une loi particulière, dite *loi de Poisson*.

Considérons la variable aléatoire discrète X pouvant prendre seulement des valeurs entières non négatives:

$$0, 1, 2, \ldots, m, \ldots,$$

la suite de ces valeurs n'ayant pas de limite théorique.

On dit que la variable aléatoire X est répartie suivant la loi de Poisson si la probabilité qu'elle prenne la valeur m est donnée par la formule:

$$P_m = \frac{a^m}{m!} e^{-a} \quad (m = 0, 1, \ldots) \tag{5.9.1}$$

où a est une certaine grandeur positive appelée *paramètre de la loi de Poisson*.

Le tableau de répartition de la variable aléatoire X dans le cas de la loi de Poisson est:

x_m	0	1	2	...	m	...
P_m	e^{-a}	$\frac{a}{1!} e^{-a}$	$\frac{a^2}{2!} e^{-a}$	...	$\frac{a^m}{m!} e^{-a}$	...

Nous allons tout d'abord vérifier que la suite des probabilités donnée par la formule (5.9.1) peut être une loi de répartition, c'est-à-dire que la somme de toutes les probabilités P_m est égale à l'unité. On a

$$\sum_{m=0}^{\infty} P_m = \sum_{m=0}^{\infty} \frac{a^m}{m!} e^{-a} = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!}.$$

Mais

$$\sum_{m=0}^{\infty} \frac{a^m}{m!} = e^a,$$

d'où

$$\sum_{m=0}^{\infty} P_m = e^{-a} e^a = 1.$$

Sur la figure 5.9.1 on peut voir les polygones de répartition de la loi de Poisson pour différentes valeurs du paramètre a. Dans la table 8 de l'annexe sont rassemblées les valeurs de P_m pour différentes valeurs de a.

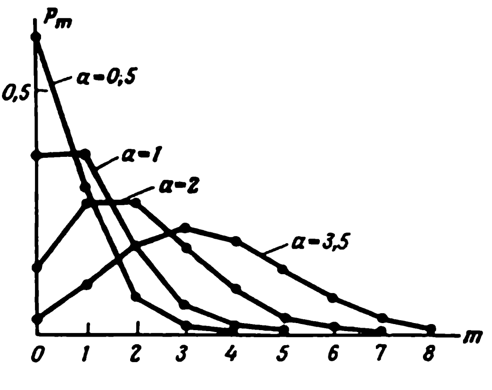


Fig. 5.9.1

Trouvons maintenant les caractéristiques essentielles, à savoir l'espérance mathématique et la variance de la loi de Poisson. Par définition de l'espérance mathématique on a:

$$m_x = M[X] = \sum_{m=0}^{\infty} mP_m = \sum_{m=0}^{\infty} m\frac{a^m}{m!}e^{-a}.$$

Le premier terme de la somme (correspondant à $m = 0$) est égal à zéro, par conséquent on peut commencer la sommation par $m = 1$:

$$m_x = \sum_{m=1}^{\infty} m\frac{a^m}{m!}e^{-a} = ae^{-a}\sum_{m=1}^{\infty}\frac{ma^{m-1}}{m!} = ae^{-a}\sum_{m=1}^{\infty}\frac{a^{m-1}}{(m-1)!}.$$

En introduisant la désignation $m - 1 = k$, on a:

$$m_x = ae^{-a}\sum_{k=0}^{\infty}\frac{a^k}{k!} = ae^{-a}e^{a} = a. \qquad (5.9.2)$$

Ainsi le paramètre a n'est rien d'autre que l'espérance mathématique de X.

Pour déterminer la variance il y a lieu de trouver préalablement le moment initial deux de X:

$$\alpha_2 = \sum_{m=0}^{\infty} m^2 \frac{a^m}{m!} e^{-a} = a \sum_{m=1}^{\infty} m \frac{a^{m-1}}{(m-1)!} e^{-a} =$$

$$= a \sum_{m=1}^{\infty} [(m-1)+1] \frac{a^{m-1}}{(m-1)!} e^{-a} =$$

$$= a \left[\sum_{m=1}^{\infty} (m-1) \frac{a^{m-1}}{(m-1)!} e^{-a} + \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} \right].$$

Nous avons montré plus haut que:

$$\sum_{m=1}^{\infty} (m-1) \frac{a^{m-1}}{(m-1)!} e^{-a} = \sum_{k=0}^{\infty} k \frac{a^k}{k!} e^{-a} = a;$$

de plus:

$$\sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} e^{-a} = e^{-a} e^{a} = 1,$$

par conséquent:

$$\alpha_2 = a(a+1)$$

et la variance de X:

$$D_x = \alpha_2 - m_x^2 = a^2 + a - a^2 = a. \qquad (5.9.3)$$

Ainsi, *la variance d'une variable aléatoire suivant la loi de Poisson est égale à son espérance mathématique a.*

Cette propriété de la loi de Poisson est souvent utilisée dans les applications pratiques quand il y a lieu d'établir la vraisemblance de l'hypothèse suivant laquelle la variable aléatoire X suit la loi de Poisson. A cet effet on détermine à partir des données expérimentales l'espérance mathématique et la variance statistiques de X *). La proximité de leurs valeurs peut témoigner en faveur de la loi de Poisson; si ces caractéristiques sont nettement différentes l'hypothèse n'est pas confirmée.

Déterminons pour une variable aléatoire X répartie suivant la loi de Poisson la probabilité pour que sa valeur ne soit pas inférieure à k. Soit R_k cette probabilité:

$$R_k = P(X \geqslant k).$$

*) Pour les méthodes expérimentales de détermination de ces caractéristiques voir chapitres 7 et 14.

Il est clair que la probabilité R_k peut être calculée comme la somme:

$$R_k = P_k + P_{k+1} + \ldots = \sum_{m=k}^{\infty} P_m.$$

Il est cependant plus simple de la trouver à partir de la probabilité de l'événement contraire:

$$R_k = 1 - (P_0 + P_1 + \ldots + P_{k-1}) = 1 - \sum_{m=0}^{k-1} P_m. \qquad (5.9.4)$$

En particulier, la probabilité pour que X soit positive est donnée par la formule:

$$R_1 = 1 - P_0 = 1 - e^{a}. \qquad (5.9.5)$$

Nous avons déjà mentionné que dans de nombreuses applications on rencontre la loi de Poisson. Nous allons étudier un problème de ce genre.

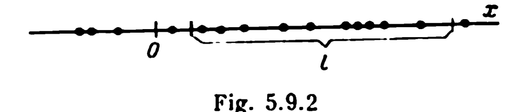


Fig. 5.9.2

Imaginons des points disposés au hasard sur l'axe des abscisses (fig. 5.9.2). Supposons que la répartition des points satisfasse aux conditions suivantes:

1. La probabilité pour un intervalle l de contenir un nombre quelconque de points dépend seulement de la longueur de cet intervalle, mais non de sa position sur l'axe des abscisses. En d'autres termes, les points sont répartis sur l'axe des abscisses avec la même densité moyenne que nous notons λ (c'est l'espérance mathématique du nombre de points par unité de longueur).

2. Les points sont répartis sur l'axe des abscisses indépendamment les uns des autres, c'est-à-dire que la probabilité pour qu'un nombre quelconque de points se trouve dans un intervalle donné ne dépend pas du nombre de points tombés dans tout autre intervalle ne se recouvrant pas avec l'intervalle considéré.

3. La probabilité pour que deux points ou plus tombent dans un petit intervalle Δx est négligeable par rapport à la probabilité d'y trouver un seul point (cette condition signifie que deux points ou plus ne peuvent pratiquement pas coïncider).

Séparons sur l'axe des abscisses un certain intervalle de longueur l et considérons la variable aléatoire discrète X associée au nombre de points se trouvant dans cet intervalle. Les valeurs possibles de cette variable sont

$$0,\ 1,\ 2,\ \ldots,\ m,\ \ldots \qquad (5.9.6)$$

Les points tombant dans l'intervalle indépendamment les uns des autres, il n'est théoriquement pas exclu que leur nombre soit infiniment grand, c'est-à-dire que la suite (5.9.6) soit sans fin.

Montrons que X suit la loi de Poisson. A cet effet calculons la probabilité P_m pour que dans l'intervalle l il y ait exactement m points.

Commençons par résoudre un problème plus simple. Considérons sur l'axe Ox un petit intervalle Δx et calculons la probabilité d'y trouver au moins un point. Nous allons raisonner comme suit. L'espérance mathématique du nombre de points se trouvant dans cet intervalle est de toute évidence égale à $\lambda\Delta x$ (car par unité de longueur il y a en moyenne λ points). En vertu de la condition 3 pour le petit intervalle Δx on peut négliger la probabilité d'y trouver plus d'un point. Ainsi, l'espérance mathématique $\lambda\Delta x$ du nombre de points se trouvant dans l'intervalle Δx sera approximativement égale à la probabilité pour qu'un point (au moins un point) s'y trouve.

Ainsi, à des infiniment petits d'ordre supérieur près, pour $\Delta x \to 0$ on peut considérer que la probabilité pour que, dans l'intervalle Δx, on ait un (au moins un) point est égale à $\lambda\Delta x$, et la probabilité pour qu'aucun point ne s'y trouve est égale à $1 - \lambda\Delta x$.

C'est ce que nous allons utiliser pour calculer la probabilité P_m pour que m points tombent dans l'intervalle l. Divisons l'intervalle l en n parties égales de longueur $\Delta x = \frac{l}{n}$. Nous dirons que l'intervalle élémentaire Δx est « vide » s'il ne contient aucun point et qu'il est « occupé » s'il en contient au moins un. Conformément à ce qui a été démontré ci-dessus, la probabilité pour que l'intervalle Δx soit « occupé » est approximativement égale à $\lambda\Delta x = \frac{\lambda l}{n}$; la probabilité pour qu'il soit « vide » est égale à $1 - \frac{\lambda l}{n}$. Comme en vertu de la condition 2 les points sont répartis indépendamment parmi les intervalles disjoints, on peut considérer nos n intervalles élémentaires comme n expériences indépendantes, dans chacune desquelles un intervalle peut être « occupé » avec une probabilité $p = \frac{\lambda l}{n}$. Cherchons la probabilité pour que, des n intervalles, soient occupés exactement m. En vertu du théorème des expériences répétées cette probabilité vaut:

$$C_n^m \left(\frac{\lambda l}{n}\right)^m \left(1 - \frac{\lambda l}{n}\right)^{n-m}$$

ou, en introduisant la désignation $\lambda l = a$:

$$C_n^m \left(\frac{a}{n}\right)^m \left(1 - \frac{a}{n}\right)^{n-m}. \tag{5.9.7}$$

Pour n suffisamment grand cette probabilité est approximativement égale à la probabilité pour que exactement m points se trouvent dans l'intervalle l, car la probabilité de trouver dans l'intervalle Δx plus d'un point est négligeable. Pour calculer la valeur exacte de P_m il faut passer dans l'expression (5.9.7) à la limite pour $n \to \infty$:

$$P_m = \lim_{n\to\infty} C_n^m \left(\frac{a}{n}\right)^m \left(1-\frac{a}{n}\right)^{n-m}. \tag{5.9.8}$$

Faisons la transformation:

$$C_n^m \left(\frac{a}{n}\right)^m \left(1-\frac{a}{n}\right)^{n-m} = \frac{n(n-1)\dots(n-m+1)}{m!}\frac{a^m}{n^m}\frac{\left(1-\frac{a}{n}\right)^n}{\left(1-\frac{a}{n}\right)^m} =$$

$$= \frac{n(n-1)\dots(n-m+1)}{n^m}\frac{a^m}{m!}\frac{\left(1-\frac{a}{n}\right)^n}{\left(1-\frac{a}{n}\right)^m}. \tag{5.9.9}$$

La première fraction et le dénominateur de la dernière fraction dans l'expression (5.9.9) tendent vers l'unité pour $n \to \infty$. L'expression $\frac{a^m}{m!}$ ne dépend pas de n. Le numérateur de la dernière fraction peut s'écrire:

$$\left(1-\frac{a}{n}\right)^n = \left[\left(1-\frac{a}{n}\right)^{\frac{n}{a}}\right]^a. \tag{5.9.10}$$

Pour $n \to \infty$ $\frac{n}{a} \to \infty$ et l'expression (5.9.10) tend vers e^{-a}. Nous avons démontré ainsi que la probabilite pour que exactement m points tombent dans l'intervalle l est donnée par l'expression:

$$P_m = \frac{a^m}{m!} e^{-a},$$

où $a = \lambda l$, c'est-à-dire que la variable X suit une loi de Poisson de paramètre $a = \lambda l$.

Remarquons que la signification de a est le nombre moyen de points tombant dans l'intervalle de longueur l.

La grandeur R_1 (probabilité pour que X soit positive) exprime la *probabilité de trouver dans l'intervalle l au moins un point*:

$$R_1 = 1 - e^{-a}. \tag{5.9.11}$$

On voit ainsi que la répartition de Poisson apparaît là où des points (ou d'autres éléments) quelconques occupent au hasard une certaine position, indépendamment les uns des autres, et que l'on compte le nombre de points tombés dans une certaine région. Dans

notre cas cette région est l'intervalle l de l'axe des abscisses. On peut généraliser facilement la démonstration au cas de la répartition des points dans le plan (champ aléatoire plan) et dans l'espace (champ aléatoire spatial). Il est facile de montrer que pour les conditions suivantes:

1) les points sont statistiquement uniformément répartis dans le champ avec une densité moyenne λ;

2) dans les régions disjointes les points tombent indépendamment les uns des autres;

3) les points apparaissent un à un et non par couples, triplets, etc., le nombre de points X tombés dans une région quelconque D (plane ou spatiale) suit une loi de Poisson

$$P_m = \frac{a^m}{m!} e^{-a} \qquad (m = 0, 1, 2, \ldots),$$

où a est le nombre moyen de points tombés dans la région D.

Dans le cas d'un plan on a:

$$a = S_D \cdot \lambda,$$

où S_D est l'aire de la région D;

pour le cas spatial:

$$a = V_D \cdot \lambda,$$

où V_D est le volume de la région D.

Notons que la condition 1) de densité constante importe peu pour que le nombre de points tombés dans une région quelconque suive une loi de Poisson, si seulement les deux autres conditions sont remplies; dans ce cas le paramètre a s'obtient différemment. Pour l'obtenir il faut, au lieu de multiplier la densité λ par la longueur, l'aire ou le volume de la région, intégrer la densité variable sur l'intervalle, l'aire ou le volume (pour plus de détails voir § 19.4).

La dispersion aléatoire de points sur une droite, dans un plan ou dans un volume n'est pas le cas unique où la loi de Poisson peut se manifester. On peut montrer, par exemple, que la loi de Poisson est la loi limite pour la répartition binomiale

$$P_{m,n} = C_n^m p^m (1-p)^{n-m}, \tag{5.9.12}$$

si on fait simultanément tendre le nombre d'expériences n vers l'infini et la probabilité p vers zéro, leur produit ayant une valeur constante

$$np = a. \tag{5.9.13}$$

En effet, exprimons cette tendance de la loi binomiale:

$$\lim_{n\to\infty} C_n^m p^m (1-p)^{n-m} = \frac{a^m}{m!} e^{-a}; \tag{5.9.14}$$

mais en vertu de la condition (5.9.13)

$$p=\frac{a}{n}, \tag{5.9.15}$$

alors en substituant (5.9.15) dans (5.9.14) nous obtenons l'égalité.

$$\lim_{n\to\infty} C_n^m\left(\frac{a}{n}\right)^m\left(1-\frac{a}{n}\right)^{n-m}=\frac{a^m}{m!}e^{-a} \tag{5.9.16}$$

que nous venons de démontrer à une autre fin.

Cette propriété limite de la loi binomiale est souvent utilisée dans les applications pratiques. Supposons qu'on effectue un grand nombre d'expériences indépendantes n dans chacune desquelles l'événement A a une probabilité infime p d'être réalisé. Pour calculer la probabilité $P_{m,n}$ de réalisation de l'événement A exactement m fois on peut se servir de la formule approchée

$$P_{m,n}\approx\frac{(np)^m}{m!}e^{-np}, \tag{5.9.17}$$

où $np = a$ est le paramètre de la loi de Poisson remplaçant approximativement la répartition binomiale.

C'est à cause de cette propriété de la loi de Poisson d'exprimer la répartition binomiale pour un grand nombre d'expériences et une probabilité infime de l'événement qu'elle est parfois appelée *loi des événements rares*.

Considérons quelques exemples où l'on rencontre la répartition de Poisson.

Exemple 1. Une centrale téléphonique reçoit des appels à raison de K appels par heure en moyenne. Supposant que le nombre d'appels pendant un intervalle de temps quelconque suit une loi de Poisson, trouver la probabilité pour que durant deux minutes la centrale reçoive exactement trois appels.

Solution. Le nombre moyen d'appels en deux minutes est:

$$a=\frac{2K}{60}=\frac{K}{30}.$$

La formule (5.9.1) donne la probabilité de trois appels:

$$P_3=\frac{\left(\frac{K}{30}\right)^3}{1\cdot 2\cdot 3}e^{-\frac{K}{30}}.$$

Exemple 2. Dans les conditions de l'exemple précédent trouver la probabilité pour qu'en deux minutes la centrale reçoive au moins un appel.

Solution. La formule (5.9.4) donne:

$$R_1=1-e^{-a}=1-e^{-K/30}.$$

Exemple 3. Dans les mêmes conditions trouver la probabilité pour qu'en deux minutes la centrale reçoive au moins trois appels.

Solution. La formule (5.9.4) donne:

$$R_3=1-(P_0+P_1+P_2)=1-e^{-K/30}\times\left[1+\frac{K}{30}+\frac{1}{2}\left(\frac{K}{30}\right)^2\right].$$

E x e m p l e 4. Le fil d'un métier à tisser se casse en moyenne 0,375 fois par heure de fonctionnement du métier. Trouver la probabilité pour que durant huit heures de travail le nombre X de cassures du fil se trouve entre 2 et 4 $(2 \leqslant X \leqslant 4)$.

S o l u t i o n. Il est évident que l'on a:

$$a = 0{,}375 \cdot 8 = 3,$$

alors:

$$P\,(2 \leqslant X \leqslant 4) = P_2 + P_3 + P_4.$$

La table 8 de l'annexe donne pour $a = 3$:

$$P_2 = 0{,}224; \quad P_3 = 0{,}224; \quad P_4 = 0{,}168,$$
$$P\,(2 \leqslant X \leqslant 4) = 0{,}616.$$

E x e m p l e 5. Une cathode chaude émet par unité de temps en moyenne $q\,(t)$ électrons, où t est le temps écoulé à partir du début de l'expérience. Trouver la probabilité que durant un temps τ commençant à l'instant t_0 la cathode émettra exactement m électrons.

S o l u t i o n. Cherchons le nombre d'électrons a émis par la cathode durant le temps τ. On a:

$$a = \int\limits_{t_0}^{t_0+\tau} q\,(t)\,dt.$$

Connaissant a, calculons la probabilité cherchée:

$$P_m = \frac{a^m}{m} e^{-a}.$$

E x e m p l e 6. Des pièces sont découpées dans une feuille de métal ayant des défauts ponctuels. Les défauts forment un champ satisfaisant aux conditions 1), 2), 3) de la page 107, de densité 0,03 point par cm². Si dans une pièce il y a au moins un défaut ponctuel, la pièce est rejetée. La surface de la pièce $S = 10$ cm². Trouver la probabilité pour une pièce d'être rejetée.

S o l u t i o n. On a $a = \lambda S = 0{,}03 \cdot 10 = 0{,}3$. La formule (5.9.5) donne la probabilité pour que dans les limites d'une pièce il y ait au moins un défaut ponctuel:

$$R_1 = 1 - e^{-0{,}3} = 1 - 0{,}741 = 0{,}259.$$

E x e m p l e 7. La densité moyenne des microbes nocifs dans un mètre cube d'air est égale à 100. On prend un échantillon de 2 dm³ d'air. Trouver la probabilité pour que dans ce volume il y ait au moins un microbe.

S o l u t i o n. Supposant que le nombre de microbes contenus dans un volume suive une loi de Poisson on a:

$$a = 0{,}2; \qquad R_1 = 1 - e^{-0{,}2} \approx 1 - 0{,}819 \approx 0{,}18.$$

E x e m p l e 8. Un dispositif technique se compose de $n = 50$ éléments pouvant chacun, durant le temps de fonctionnement t, tomber en panne (les uns indépendamment des autres) avec une probabilité égale à 0,04. En se basant sur la propriété limite de la répartition binomiale (formule (5.9.17)) trouver la probabilité pour que durant le temps t aucun élément ne soit en panne, ou qu'un, deux... éléments soient en panne.

S o l u t i o n. On a $a = np = 50 \cdot 0{,}04 = 2$. La table 8 de l'annexe donne pour $a = 2$:

$$P_0 \approx 0{,}135; \quad P_1 \approx 0{,}271; \quad P_2 \approx 0{,}271; \quad P_3 \approx 0{,}180;$$
$$P_4 \approx 0{,}090; \quad P_5 \approx 0{,}036; \quad P_6 \approx 0{,}012; \quad P_7 \approx 0{,}004;$$
$$P_8 \approx 0{,}001; \quad P_9 \approx P_{10} \approx \ldots \approx 0.$$

CHAPITRE 6

LOI NORMALE

6.1. La loi normale et ses paramètres

La loi de répartition normale (souvent appelée loi de Gauss) joue un rôle particulièrement important dans la théorie des probabilités et dans les applications pratiques en occupant une place de choix parmi les lois de répartition. La particularité fondamentale de la loi normale la distinguant des autres lois est que c'est une *loi limite* vers laquelle tendent les autres lois pour des conditions se rencontrant fréquemment dans les applications pratiques.

On peut montrer que la somme d'un nombre suffisamment grand de variables aléatoires indépendantes (ou faiblement liées) suivant des lois quelconques, pour des restrictions assez faibles, tend approximativement vers une loi normale et ceci avec d'autant plus de précision que le nombre de termes est important. La majorité des variables aléatoires que l'on rencontre dans la pratique, comme par exemple les erreurs de mesure, les erreurs de tir, etc. peuvent être considérées comme des sommes d'un nombre relativement important de termes, erreurs élémentaires, dues chacune à une cause différente indépendante des autres. Quelle que soit la loi des erreurs élémentaires, les particularités de ces répartitions n'apparaissent pas dans la somme d'un grand nombre de celles-ci, la somme suivant une loi voisine de la loi normale. La seule limitation imposée aux erreurs sommées est de jouer dans la somme un rôle relativement peu important. Si cette condition ne se trouve pas remplie et que par exemple l'une des erreurs aléatoires prévaut nettement sur les autres, celle-ci détermine en gros traits la loi de répartition de la somme.

Les théorèmes établissant que la loi normale est la loi limite d'une somme de termes aléatoires indépendants uniformément petits seront étudiés dans le chapitre 13.

La loi normale est caractérisée par la densité de probabilité de la forme:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}. \qquad (6.1.1)$$

La courbe de la densité de la loi normale a la forme d'une cloche symétrique (fig. 6.1.1). L'ordonnée maximale de la courbe, égale

à $\frac{1}{\sigma\sqrt{2\pi}}$, correspond au point $x = m$; au fur et à mesure que l'on s'éloigne du point m la densité de probabilité diminue et pour $x \to \pm\infty$ elle tend asymptotiquement vers l'axe des abscisses.

Donnons une interprétation aux paramètres m et σ entrant dans l'expression de la loi normale (6.1.1). Nous allons démontrer que m n'est rien d'autre que l'*espérance mathématique*, et σ, l'*écart quadratique moyen* de X. A cet effet nous calculons les caractéristiques numériques essentielles de X, c'est-à-dire son espérance mathématique et sa variance:

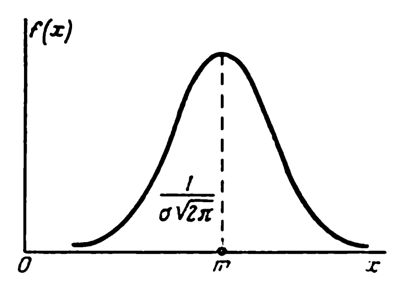


Fig. 6.1.1

$$M[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx.$$

En introduisant le changement de variable

$$\frac{x-m}{\sigma\sqrt{2}} = t,$$

on obtient:

$$M[X] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}\,t + m)\, e^{-t^2}\, dt = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} t e^{-t^2}\, dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}\, dt. \qquad (6.1.2)$$

Il est facile de voir que la première des deux intégrales dans (6.1.2) est nulle; la seconde est l'intégrale d'Euler-Poisson:

$$\int_{-\infty}^{\infty} e^{-t^2}\, dt = 2\int_{0}^{\infty} e^{-t^2}\, dt = \sqrt{\pi}. \qquad (6.1.3)$$

Par conséquent:

$$M[X] = m,$$

c'est-à-dire que le paramètre m est l'espérance mathématique de la variable X. On appelle souvent ce paramètre, surtout dans les problèmes de tir, *centre de dispersion*.

Calculons la variance de X:

$$D[X]=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}dx.$$

En changeant de nouveau de variable:

$$\frac{x-m}{\sigma\sqrt{\pi}}=t,$$

on a:

$$D[X]=\frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^2 e^{-t^2}dt.$$

L'intégration par parties donne:

$$D[X]=\frac{\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}t\cdot 2te^{-t^2}dt=\frac{\sigma^2}{\sqrt{\pi}}\left\{-te^{-t^2}\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}e^{-t^2}dt\right\}.$$

Le premier terme dans l'accolade est égal à zéro (car pour $t\to\infty$ e^{-t^2} décroît plus rapidement que ne croît aucune puissance de t),

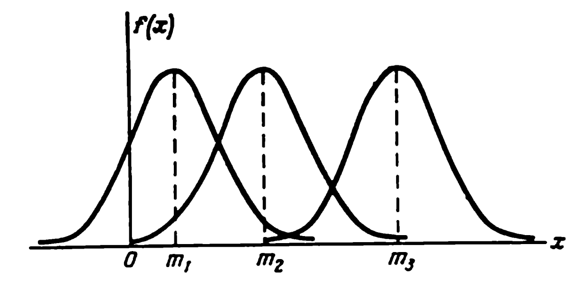


Fig. 6.1.2

le second terme dans la formule (6.1.3) est égal à $\sqrt{\pi}$, d'où:

$$D[X]=\sigma^2.$$

Par conséquent le paramètre σ dans la formule (6.1.1) n est rien d'autre que l'écart quadratique moyen de la variable X.

Arrêtons-nous sur la signification physique des paramètres m et σ. La formule (6.1.1) montre que le centre de dispersion m est le centre de symétrie de la répartition. En effet lorsque la différence $(x-m)$ change de signe, l'expression (6.1.1) reste inchangée. Si m change de valeur, la courbe de densité se déplace suivant l'axe des abscisses sans changer de forme (fig. 6.1.2). Le centre de dispersion caractérise la position de la répartition sur l'axe des abscisses.

La dimension du centre de dispersion est celle de la variable aléatoire X.

Le paramètre σ caractérise non pas la position mais la forme de la courbe de densité, c'est une caractéristique de la dispersion. L'ordonnée maximale de la courbe de répartition est inversement proportionnelle à σ; lorsque σ augmente l'ordonnée maximale diminue. L'aire limitée par la courbe de densité devant toujours être égale à l'unité, avec augmentation de σ la courbe devient plus plate et s'étale en abscisses; au contraire lorsque σ diminue, la courbe s'allonge vers le haut en se rétrécissant des côtés, devenant de plus en plus pointue. Sur la figure 6.1.3 on peut voir trois courbes de lois

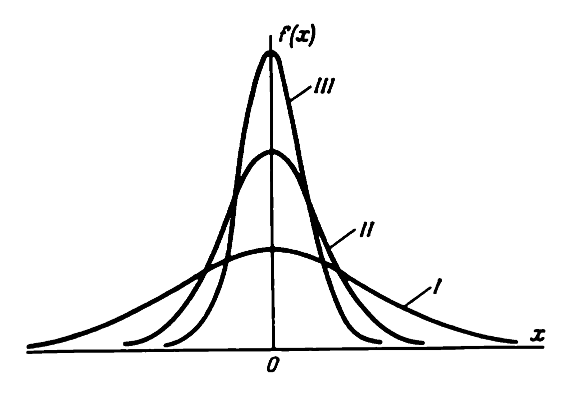


Fig. 6.1.3

normales (*I*, *II*, *III*) pour $m = 0$; $\sigma_1 > \sigma_2 > \sigma_3$. Le changement du paramètre σ équivaut à une augmentation de l'échelle suivant l'un des axes et à une même diminution suivant l'autre axe.

Il est évident que le paramètre σ a même dimension que la variable aléatoire X.

Dans certains cours de théorie des probabilités pour caractériser la dispersion de la loi normale on utilise non pas l'écart quadratique moyen mais une grandeur qui lui est inversement proportionnelle, appelée *mesure de précision* et définie par

$$h = \frac{1}{\sigma\sqrt{2}}.$$

La dimension de la mesure de précision est l'inverse de celle de X.

Le terme « mesure de précision » a été emprunté à la théorie des erreurs. En utilisant la mesure de précision h on peut écrire comme suit la loi normale:

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2}.$$

6.2. Moments de la loi normale

Nous avons montré ci-dessus que l'espérance mathématique d'une variable aléatoire suivant la loi normale (6.1.1) est égale à m, et l'écart quadratique moyen à σ.

Nous allons obtenir les formules générales pour les moments centrés d'ordre quelconque.

Par définition:

$$\mu_s=\int_{-\infty}^{\infty}(x-m)^s f(x)\,dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-m)^s e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx.$$

Après le changement de variable

$$\frac{x-m}{\sigma\sqrt{2}}=t$$

on a:

$$\mu_s=\frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^s e^{-t^2}\,dt. \tag{6.2.1}$$

Intégrons l'expression (6.2.1) par parties, il vient:

$$\mu_s=\frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^{s-1}te^{-t^2}\,dt=$$

$$=\frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}}\left\{-\frac{1}{2}e^{-t^2}t^{s-1}\Big|_{-\infty}^{\infty}+\frac{s-1}{2}\int_{-\infty}^{\infty}t^{s-2}e^{-t^2}\,dt\right\}.$$

Le premier terme dans l'accolade étant nul, on a:

$$\mu_s=\frac{(s-1)(\sigma\sqrt{2})^s}{2\sqrt{\pi}}\int_{-\infty}^{\infty}t^{s-2}e^{-t^2}\,dt. \tag{6.2.2}$$

A partir de (6.2.1) on obtient pour μ_{s-2} l'expression suivante:

$$\mu_{s-2}=\frac{(\sigma\sqrt{2})^{s-2}}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^{s-2}e^{-t^2}\,dt. \tag{6.2.3}$$

En comparant les seconds membres des formules (6.2.2) et (6.2.3) on voit qu'ils ne diffèrent que par le facteur $(s-1)\,\sigma^2$, par conséquent:

$$\mu_s=(s-1)\,\sigma^2\mu_{s-2}. \tag{6.2.4}$$

La formule (6.2.4) est une formule de récurrence permettant d'exprimer les moments d'ordre supérieur en fonction des moments

d'ordre inférieur. Compte tenu de $\mu_0 = 1$ *) et $\mu_1 = 0$, cette formule permet de calculer les moments centrés d'ordre quelconque. Comme $\mu_1 = 0$, il est immédiat que tous les moments impairs de la loi normale sont nuls. Ceci découle d'ailleurs de la symétrie de la loi normale.

La formule (6.2.4) pour s pair donne les expressions suivantes pour les moments:

$$\mu_2 = \sigma^2;$$
$$\mu_4 = 3\sigma^4;$$
$$\mu_6 = 15\sigma^6, \text{ etc,}$$

La formule générale donnant le moment d'ordre s pour s pair quelconque est:

$$\mu_s = (s-1)!!\,\sigma^s,$$

où le symbole $(s-1)!!$ désigne le produit de tous les nombres impairs de 1 à $s-1$.

Comme pour la loi normale on a $\mu_3 = 0$, l'asymétrie est également nulle:

$$Sk = \frac{\mu_3}{\sigma^3} = 0.$$

L'expression du moment quatre:

$$\mu_4 = 3\sigma^4$$

permet de trouver:

$$Ex = \frac{\mu_4}{\sigma^4} - 3 = 0,$$

autrement dit, l'aplatissement de la loi normale est nul. C'est tout naturel car le coefficient d'aplatissement caractérise justement la pente d'une loi de répartition quelconque par rapport à la loi normale.

6.3. Probabilité pour une variable aléatoire normale de tomber dans un intervalle donné. Fonction de répartition normale

Dans de nombreux problèmes traitant des variables aléatoires normales souvent il y a lieu de trouver la probabilité pour une variable aléatoire X, répartie suivant une loi normale de paramètres m, σ, de tomber dans l'intervalle α, β. Pour calculer cette probabilité nous allons utiliser la formule générale:

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha), \qquad (6.3.1)$$

où $F(x)$ est la fonction de répartition de X.

*) Le moment d'ordre zéro de toute variable aléatoire est égal à l'unité en tant qu'espérance mathématique de la puissance zéro de cette variable.

Nous allons trouver la fonction de répartition $F(x)$ de la variable aléatoire X répartie suivant une loi normale de paramètres m, σ. La densité de probabilité de X est:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad (6.3.2)$$

d'où la fonction de répartition:

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx. \qquad (6.3.3)$$

Introduisons dans l'intégrale (6.3.3) le changement de variable:

$$\frac{x-m}{\sigma}=t,$$

on obtient alors:

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}}\,dt. \qquad (6.3.4)$$

L'intégrale (6.3.4) ne peut s'exprimer par des fonctions élémentaires mais elle peut être calculée à l'aide d'une fonction spéciale qui est une intégrale définie de e^{-t^2} ou de $e^{-\frac{t^2}{2}}$ (appelée *intégrale des probabilités*) laquelle est tabulée. Il existe un certain nombre de fonctions de ce genre, par exemple:

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_{0}^{x} e^{-t^2}\,dt;$$

$$\Phi_1(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-\frac{t^2}{2}}\,dt, \text{ etc.}$$

Parmi les fonctions de cette espèce nous choisirons

$$\Phi^*(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \qquad (6.3.5)$$

Il est facile de voir que cette expression n'est rien d'autre que la loi de répartition normale de paramètres $m=0$, $\sigma=1$.

Convenons d'appeler la fonction $\Phi^*(x)$ *fonction de répartition normale*. On peut trouver dans l'annexe (table 1) les valeurs tabulées de la fonction $\Phi^*(x)$ *).

*) Pour faciliter l'interpolation, dans les tables, à côté des valeurs de la fonction on donne son accroissement pour un pas Δ.

Exprimons la fonction de répartition (6.3.3) à l'aide de la fonction de répartition normale $\Phi^*(x)$. Il est évident que

$$F(x) = \Phi^*\left(\frac{x-m}{\sigma}\right) \tag{6.3.6}$$

Nous pouvons maintenant trouver la probabilité pour la variable aléatoire X d'atteindre l'intervalle compris entre α et β. En vertu de la formule (6.3.1) on a:

$$P(\alpha < X < \beta) = \Phi^*\left(\frac{\beta-m}{\sigma}\right) - \Phi^*\left(\frac{\alpha-m}{\sigma}\right). \tag{6.3.7}$$

Nous avons exprimé la probabilité pour la variable aléatoire X répartie suivant une loi normale de paramètres quelconques de tomber dans l'intervalle donné, à l'aide de la fonction de répartition normale $\Phi^*(x)$ correspondant à la loi normale simple de paramètres 0 et 1. Notons que les arguments de la fonction Φ^* dans la formule (6.3.7) ont une interprétation très simple. En effet, $\frac{\beta-m}{\sigma}$ et $\frac{\alpha-m}{\sigma}$ sont les distances des extrémités droite et gauche respectivement jusqu'au centre de dispersion, exprimées en écarts quadratiques moyens. La distance est considérée positive si l'extrémité de l'intervalle est à droite du centre de dispersion et négative si elle est à gauche.

Comme toute fonction de répartition, la fonction $\Phi^*(x)$ est douée des propriétés suivantes:

1. $\Phi^*(-\infty) = 0$.
2. $\Phi^*(+\infty) = 1$.
3. $\Phi^*(x)$ est une fonction non décroissante.

En effet, en vertu de la symétrie de la loi normale de paramètres $m = 0$, $\sigma = 1$ par rapport à l'origine des coordonnées, on a:

$$\Phi^*(-x) = 1 - \Phi^*(x). \tag{6.3.8}$$

En utilisant cette propriété on aurait pu se limiter dans les tables de la fonction $\Phi^*(x)$ uniquement aux valeurs positives de l'argument, mais pour épargner une opération supplémentaire (retranchement de l'unité), dans la table 1 de l'annexe on donne les valeurs de $\Phi^*(x)$ tant pour les arguments positifs que négatifs.

Dans les applications pratiques souvent il y a lieu de calculer la probabilité pour une variable normale d'atteindre le domaine symétrique par rapport au centre de dispersion m. Soit un tel domaine de longueur $2l$ (fig. 6.3.1), la formule (6.3.7) permet de calculer la probabilité d'atteindre ce domaine, soit:

$$P(m-l < X < m+l) = \Phi^*\left(\frac{l}{\sigma}\right) - \Phi^*\left(-\frac{l}{\sigma}\right). \tag{6.3.9}$$

Vu la propriété (6.3.8) de la fonction $\Phi^*(x)$ et en simplifiant l'écriture de la formule (6.3.9), on obtient la formule donnant la

probabilité pour une variable aléatoire normale d'atteindre le domaine symétrique par rapport au centre de dispersion:

$$P(|X-m|<l)=2\Phi^*\left(\frac{l}{\sigma}\right)-1. \qquad (6.3.10)$$

Proposons nous de résoudre le problème suivant. A partir du centre de dispersion m on porte des segments successifs de longueur σ (fig. 6.3.2) et on calcule la probabilité pour la variable aléatoire X

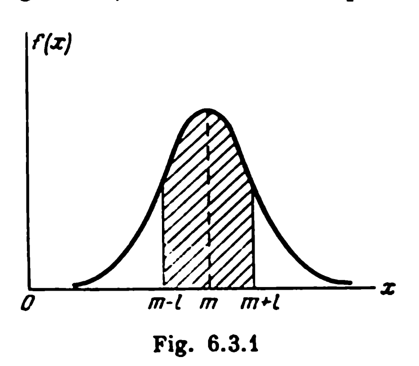


Fig. 6.3.1

Fig. 6.3.2

d'atteindre chacun d'eux. La courbe de la loi normale étant symétrique, il suffit de porter ces segments d'un seul côté.

La formule (6.3.7) donne:

$$\left.\begin{aligned}P(m<X<m+\sigma)&=\Phi^*(1)-\Phi^*(0)=\\&=0{,}8413-0{,}5000\approx 0{,}341;\\P(m+\sigma<X<m+2\sigma)&=\Phi^*(2)-\Phi^*(1)\approx 0{,}136;\\P(m+2\sigma<X<m+3\sigma)&=\Phi^*(3)-\Phi^*(2)\approx 0{,}012;\\P(m+3\sigma<X<m+4\sigma)&=\Phi^*(4)-\Phi^*(3)\approx 0{,}001.\end{aligned}\right\} \qquad (6.3.11)$$

Les résultats obtenus montrent que la probabilité de tomber dans chacun des segments suivants (cinquième, sixième, etc.) est égale à zéro à 0,001 près.

En arrondissant les trois premiers chiffres à 0,01 (à 1 % près) on obtient trois nombres faciles à retenir:

$$0{,}34;\ 0{,}14;\ 0{,}02.$$

La somme de ces trois valeurs est égale à 0,5. Ceci signifie que pour la répartition normale d'une variable aléatoire toute la dispersion (à des fractions de pour cent près) se limite à l'intervalle $m \pm 3\sigma$.

Ceci permet, connaissant l'écart quadratique moyen et l'espérance mathématique d'une variable aléatoire, d'indiquer l'intervalle de ses valeurs possibles. Cette méthode d'estimation de la gamme des valeurs possibles d'une variable aléatoire s'appelle en mathématique

statistique « loi des trois sigma ». La loi des trois sigma permet également de trouver la valeur approchée de l'écart quadratique moyen de la variable aléatoire: on prend l'écart maximal pratiquement possible de la moyenne et on le divise par trois. Il est évident que cette méthode grossière ne peut être recommandée que s'il n'y a pas d'autres méthodes plus précises pour trouver σ.

Exemple 1. Soit la variable aléatoire X normalement répartie et correspondant à l'erreur de mesure d'une certaine distance. Lors de la mesure on fait une erreur systématique par excès de 1,2 m; l'écart quadratique moyen de l'erreur de mesure est égal à 0,8 m. Trouver la probabilité pour que l'écart entre la valeur mesurée et la valeur réelle ne soit pas en valeur absolue supérieur à 1,6 m.

Solution. L'erreur de mesure est une variable aléatoire X suivant une loi normale de paramètres $m = 1{,}2$ et $\sigma = 0{,}8$. Il s'agit de trouver la probabilité pour cette variable de tomber dans l'intervalle $\alpha = -1{,}6$ à $\beta = +1{,}6$. En vertu de la formule (6.3.7) on a:

$$P(-1{,}6 < X < 1{,}6) = \Phi^*\left(\frac{1{,}6-1{,}2}{0{,}8}\right) - \Phi^*\left(\frac{-1{,}6-1{,}2}{0{,}8}\right) = \\ = \Phi^*(0{,}5) - \Phi^*(-3{,}5).$$

En utilisant les tables de la fonction $\Phi^*(x)$ (annexe table 1) ou trouve:

$$\Phi^*(0{,}5) = 0{,}6915; \quad \Phi^*(-3{,}5) = 0{,}0002.$$

d'où:

$$P(-1{,}6 < X < 1{,}6) = 0{,}6915 - 0{,}0002 = 0{,}6913 \approx 0{,}691.$$

Exemple 2. Trouver la même probabilité que dans l'exemple précédent pour le cas où il n'y a pas d'erreur systématique.

Solution. En vertu de la formule (6.3.10), posant $l = 1{,}6$, on trouve:

$$P(|X| < 1{,}6) = 2\Phi^*\left(\frac{1{,}6}{0{,}8}\right) - 1 \approx 0{,}955.$$

Exemple 3. Une usine fabrique des billes de diamètre nominal 0,8 cm; les défauts d'usinage donnent une erreur de diamètre répartic suivant une loi normale, de moyenne nulle (il n'y a pas d'erreur systématique) et d'écart quadratique moyen $\sigma = 0{,}001$ cm. Lors du contrôle sont mises à rebut toutes les billes qui passent dans une bague de diamètre 0,798 cm ainsi que les billes ne pouvant passer dans une bague de diamètre 0,802 cm. Trouver la probabilité pour qu'une bille prise au hasard soit rejetée.

Solution. Soit D le diamètre de la bille. L'événement A suivant lequel une bille est rejetée se compose de deux événements:

$$A = A_1 + A_2,$$

où l'événement A_1 correspond au cas $D < 0{,}798$, ct l'événement A_2 au cas $D > 0{,}802$. Calculons la probabilité de l'événement contraire $\bar{A}$ consistant en ce que la bille est acceptée, c'est-à-dire que le diamètre D se trouve entre 0,798 et 0,802 cm:

$$P(\bar{A}) = P(0{,}798 < D < 0{,}802) = P(|D - m_d| < 0{,}002),$$

où $m_d = 0{,}8$ est le diamètre nominal de la bille. On a

$$P(\bar{A}) = 2\Phi^*\left(\frac{0{,}002}{0{,}001}\right) - 1 = 2\Phi^*(2) - 1 = 2\cdot 0{,}9772 - 1 \approx 0{,}954,$$

d'où

$$P(A) \approx 1 - 0{,}954 = 0{,}046.$$

E x e m p l e 4. Soient une variable aléatoire X normalement répartie de centre de dispersion m (fig. 6.3.3), et un certain intervalle (α, β) de l'axe des abscisses. Quelle doit être la valeur de l'écart quadratique moyen σ de la variable aléatoire X pour que la probabilité p d'atteindre l'intervalle (α, β) soit maximale?

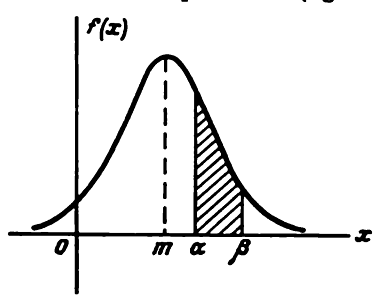


Fig. 6.3.3

S o l u t i o n. On a:

$$p=P(\alpha<X<\beta)=$$

$$=\Phi^*\left(\frac{\beta-m}{\sigma}\right)-\Phi^*\left(\frac{\alpha-m}{\sigma}\right)=\varphi(\sigma).$$

Dérivons cette fonction de σ:

$$\varphi'(\sigma)=\left[\Phi^*\left(\frac{\beta-m}{\sigma}\right)\right]'_\sigma-\left[\Phi^*\left(\frac{\alpha-m}{\sigma}\right)\right]'_\sigma,$$

mais

$$\Phi^*(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt.$$

En utilisant la règle de dérivation d'une intégrale par rapport à la variable figurant dans sa limite, on obtient:

$$\left[\Phi^*\left(\frac{\beta-m}{\sigma}\right)\right]'_\sigma=\left\{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\beta-m}{\sigma}} e^{-\frac{t^2}{2}}\,dt\right\}_\sigma=$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\beta-m)^2}{2\sigma^2}}\left(-\frac{\beta-m}{\sigma^2}\right)=-\frac{\beta-m}{\sigma^2\sqrt{2\pi}}e^{-\frac{(\beta-m)^2}{2\sigma^2}}$$

D'une manière analogue:

$$\left[\Phi^*\left(\frac{\alpha-m}{\sigma}\right)\right]'_\sigma=-\frac{\alpha-m}{\sigma^2\sqrt{2\pi}}e^{-\frac{(\alpha-m)^2}{2\sigma^2}}.$$

Pour trouver l'extremum on pose:

$$\varphi'(\sigma)=\frac{1}{\sigma^2\sqrt{2\pi}}\left\{(\alpha-m)e^{-\frac{(\alpha-m)^2}{2\sigma^2}}-(\beta-m)e^{-\frac{(\beta-m)^2}{2\sigma^2}}\right\}. \quad (6.3.12)$$

Pour $\sigma=\infty$ cette expression s'annule et la probabilité p devient maximale. Le maximum de p est donné par la condition:

$$(\alpha-m)e^{-\frac{(\alpha-m)^2}{2\sigma^2}}-(\beta-m)e^{-\frac{(\beta-m)^2}{2\sigma^2}}=0. \quad (6.3.13)$$

On peut résoudre l'équation (6.3.13) soit numériquement, soit graphiquement.

6.4. Ecart probable

Dans certaines applications de la théorie des probabilités (en particulier dans la théorie du tir) en plus de l'écart quadratique moyen on utilise encore une caractéristique de dispersion, à savoir l'écart *probable*, ou l'écart *médian*. L'écart probable est généralement désigné par la lettre E (parfois B).

On appelle *écart probable* (*médian*) de la variable aléatoire X normale, la moitié de la longueur de l'intervalle symétrique par rapport au centre de dispersion tel que la probabilité d'y tomber soit égale à 1/2.

Sur la figure 6.4.1 on a donné l'interprétation géométrique de l'écart probable. L'écart probable E est la moitié de la longueur du segment de l'axe des abscisses, symétrique par rapport au point m et servant de base à la figure d'aire égale à la moitié de l'aire de la courbe de répartition.

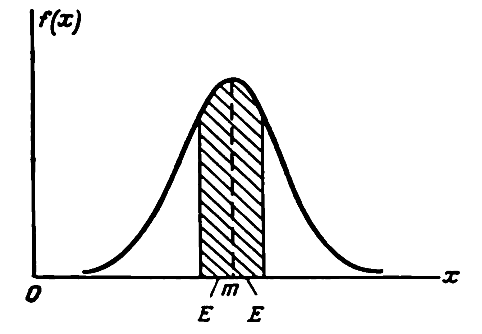


Fig. 6.4.1

Nous allons préciser la signification du terme « écart médian » ou « erreur médiane » qui est souvent utilisé dans l'artillerie au lieu de l'écart probable.

Considérons une variable aléatoire X répartie suivant une loi normale. La probabilité pour qu'elle s'écarte du centre de dispersion m d'une valeur inférieure à E est, par définition de l'écart probable E, égale à 1/2:

$$P(|X-m|<E)=\frac{1}{2}. \qquad (6.4.1)$$

La probabilité pour qu'elle s'écarte de m de plus de E est également égale à 1/2:

$$P(|X-m|>E)=\frac{1}{2}.$$

Ainsi, lorsque le nombre d'expériences est important, environ la moitié des valeurs de X s'écartera de m de plus de E et la moitié de moins de E. D'où le terme « erreur médiane », « écart médian ».

Il est évident que l'écart probable en tant que caractéristique de la dispersion doit dépendre directement de l'écart quadratique moyen σ. Nous allons établir cette relation. A cet effet nous calculons la probabilité de l'événement $|X-m|<E$ dans l'équation (6.4.1) à l'aide de la formule (6.3.10):

$$P(|X-m|<E)=2\Phi^*\left(\frac{E}{\sigma}\right)-1=\frac{1}{2},$$

d'où

$$\Phi^*\left(\frac{E}{\sigma}\right)=\frac{3}{4}=0,75. \qquad (6.4.2)$$

Les tables de la fonction $\Phi^*(x)$ permettent de trouver la valeur de l'argument x telle que $\Phi^*(x) = 0,75$. On trouve approximativement $x = 0,674$, donc:

$$\frac{E}{\sigma}=0,674; \quad E=0,674\sigma. \qquad (6.4.3)$$

Ainsi, connaissant σ on peut immédiatement trouver la valeur de E qui lui est proportionnelle. Souvent on écrit comme suit cette relation:

$$E=\rho\sqrt{2}\,\sigma, \qquad (6.4.4)$$

où ρ est la valeur de l'argument pour laquelle l'une des formes de l'intégrale de probabilité, appelée *fonction de Laplace* et donnée par

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt,$$

est égale à 1/2. La valeur numérique de ρ est approximativement égale à 0,477.

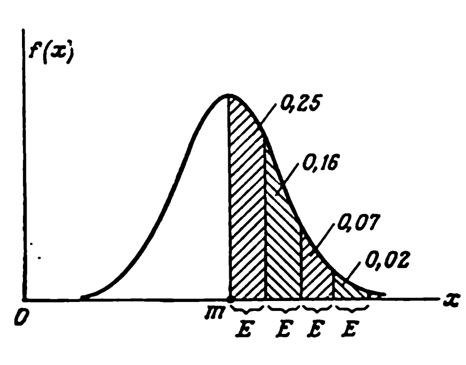


Fig. 6.4.2

A l'heure actuelle l'écart probable, en tant que caractéristique de la dispersion, cède la place à une caractéristique plus universelle σ. Cependant dans certains domaines d'application de la théorie des probabilités elle continue à s'employer par tradition.

Si pour caractériser la dispersion on prend l'écart probable E, la densité de probabilité de la répartition normale s'écrira:

$$f(x)=\frac{\rho}{E\sqrt{\pi}}e^{-\frac{\rho^2}{E^2}(x-m)^2}, \qquad (6.4.5)$$

et la probabilité pour X de se trouver entre α et β prendra la forme:

$$P(\alpha<X<\beta)=\frac{1}{2}\left[\hat{\Phi}\left(\frac{\beta-m}{E}\right)-\hat{\Phi}\left(\frac{\alpha-m}{E}\right)\right], \qquad (6.4.6)$$

où

$$\hat{\Phi}(x)=\frac{2}{\sqrt{\pi}}\int_0^{\rho x} e^{-t^2}\,dt \qquad (6.4.7)$$

est la *fonction de Laplace réduite*.

Faisons un calcul analogue à celui du paragraphe précédent pour l'écart quadratique moyen σ, c'est-à-dire, portons à partir du centre de dispersion m des segments successifs de longueur E (fig. 6.4.2) et calculons les probabilités d'atteindre ces segments à 0,01 près. On a :

$$P(m < X < m + E) \approx 0{,}25;$$
$$P(m + E < X < m + 2E) \approx 0{,}16;$$
$$P(m + 2E < X < m + 3E) \approx 0{,}07;$$
$$P(m + 3E < X < m + 4E) \approx 0{,}02.$$

On remarque qu'avec une précision de 0,01 toutes les valeurs d'une variable aléatoire normale se trouvent dans l'intervalle $m \pm 4E$.

E x e m p l e. Un avion a pour mission de larguer des colis sur une autoroute d'une largeur de 8 mètres. Le vol a lieu le long de l'autoroute, la visée suivant sa ligne moyenne; le glissement donne lieu à une erreur systématique de 2 m à droite dans la direction du vol. Les écarts probables principaux sont: dans la direction du vol $B_1 = 15$ m, latéralement $B_2 = 5$ m. Sans utiliser des tables de probabilités, ne connaissant que les nombres

25 %, 16 %, 7 %, 2 %,

estimer grossièrement la probabilité pour un colis de tomber sur l'autoroute et la probabilité pour au moins un colis des trois largués d'y tomber.

S o l u t i o n. Pour résoudre le problème il suffit de considérer une des coordonnées du point de chute, à savoir l'abscisse X dans une direction perpendiculaire à l'autoroute. Cette abscisse est répartie suivant une loi normale de centre de dispersion $m = 2$ et d'écart probable $B_2 = E = 5$ m. Portons de part et d'autre du centre de dispersion un segment de 2 m correspondant à 0,4 de l'écart probable. La probabilité d'atteindre ce segment est approximativement égale à

$$0{,}4 \cdot 25\ \% = 0{,}1.$$

A gauche du centre de dispersion la cible s'étend à 6 m. Sur ces 6 m se placent un écart probable entier (soit 5 m), auquel est associée la probabilité 25 % plus une partie (1 m) d'un autre écart probable, la probabilité associée à ce second écart étant 16 %. La probabilité d'atteindre le domaine de longueur 1 m est approximativement égale à:

$$\frac{1}{5} \cdot 16\% = 0{,}03.$$

Ainsi la probabilité de larguer le chargement sur l'autoroute est approximativement égale à:

$$0{,}1 + 0{,}25 + 0{,}03 = 0{,}38.$$

La probabilité pour un colis au moins des trois largués de tomber sur l'autoroute est égale à:

$$R_1 = 1 - (1 - 0{,}38)^3 \approx 0{,}76.$$

CHAPITRE 7

DÉTERMINATION DES LOIS DE RÉPARTITION DES VARIABLES ALÉATOIRES À PARTIR DES DONNÉES EXPÉRIMENTALES

7.1. Problèmes de base des mathématiques statistiques

Les lois mathématiques de la théorie des probabilités ne sont pas une pure abstraction dénuée de sens physique mais l'expression mathématique des lois régissant en fait les phénomènes aléatoires de la nature se produisant de nombreuses fois.

Jusqu'à présent, en parlant des lois de répartition des variables aléatoires nous n'avons pas touché à la question de l'origine et des fondements de ces lois. On peut répondre à cette question sans ambiguïté que ces lois sont établies expérimentalement; toute étude des phénomènes aléatoires par les méthodes de la théorie des probabilités est basée, directement ou indirectement, sur des données expérimentales. En introduisant les notions d'événement et de sa probabilité, de variable aléatoire, de loi de répartition et de caractéristiques numériques, la théorie des probabilités permet de déterminer théoriquement les probabilités des événements en fonction des probabilités d'autres événements, les lois de répartition et les caractéristiques des variables aléatoires en fonction des lois de répartition et des caractéristiques numériques d'autres variables aléatoires. Ces méthodes indirectes permettent de gagner du temps et d'économiser sur les frais des expériences mais n'excluent pas celles-ci. Toute étude des phénomènes aléatoires, si abstraite soit-elle, a ses racines dans l'expérience et les observations.

La statistique théorique a pour objet la recherche et la mise au point des méthodes d'enregistrement, de description et d'analyse des données statistiques expérimentales fournies par l'observation des séries de faits ou de phénomènes.

Les problèmes de la statistique concernent toujours les séries de faits aléatoires, mais peuvent prendre différentes formes suivant le problème concret à résoudre et la quantité de données dont on dispose.

Nous allons décrire brièvement certains problèmes typiques de la statistique que l'on rencontre couramment dans la pratique.

1. Détermination de la loi de répartition d'une variable aléatoire (ou d'un système de variables aléatoires) d'après les données statistiques

Nous avons déjà mentionné que les lois observées dans les séries de faits aléatoires sont d'autant plus exactes et mieux prononcées que les données statistiques dont on dispose sont abondantes. Lors du traitement d'un grand nombre de données statistiques souvent il y a lieu de déterminer les lois de répartition de telles ou telles variables aléatoires.

Théoriquement, lorsque le nombre d'expériences est suffisant. les lois propres à ces variables aléatoires peuvent se manifester d'une manière aussi précise que l'on veut. Dans la pratique on a toujours affaire à un nombre limité de données expérimentales; donc dans les résultats d'observations et de traitement est présent toujours un élément du hasard plus ou moins prononcé. Le problème qui se pose est de savoir reconnaître les propriétés stables, réellement propres au phénomène étudié et les distinguer des éléments aléatoires, apparaissant dans une certaine série de mesures et liés au nombre limité de données expérimentales.

Il est naturel d'exiger de la méthode de traitement des données expérimentales de conserver, dans la mesure du possible, les traits caractéristiques, typiques du phénomène observé et de rejeter ce qui n'est pas important, secondaire, lié à une quantité insuffisante de données expérimentales. On voit ainsi apparaître un problème typique de mathématiques statistiques, problème de *lissage* des données statistiques, de leur représentation compacte par des relations analytiques simples.

2. Vérification de la vraisemblance des hypothèses

Ce problème est étroitement lié au précédent. Comme les données statistiques dont on dispose ne sont pas, en général, suffisantes pour que les lois statistiques apparaissant ne soient pas aléatoires, ces données peuvent seulement, avec un certain degré de certitude, confirmer ou ne pas confirmer telle ou telle hypothèse. On peut par exemple se demander si les résultats des expériences sont en accord avec l'hypothèse suivant laquelle la variable aléatoire en question suit la loi de répartition $F(x)$.

Une autre question possible est de savoir si la tendance observée dans l'expérience vers une certaine loi exprime une relation réelle entre les variables aléatoires étudiées ou bien elle est due a des causes aléatoires liées à une insuffisance de données. Pour résoudre ces problèmes, des méthodes spéciales ont été élaborées en statistique.

3. Recherche des paramètres inconnus d'une loi de répartition

Il ne s'agit pas toujours lors du traitement des données statistiques de déterminer les lois de répartition des variables aléatoires étudiées. Le plus souvent ceci est lié à une quantité insuffisante de données expérimentales. Parfois on connaît d'avance le type de la loi de répartition à partir des considérations théoriques; par exemple, on peut affirmer d'avance que la variable aléatoire en question suit une loi normale. Alors le problème que doit résoudre le traitement des observations devient plus restreint; on se propose de déterminer seulement certains paramètres (caractéristiques numériques) d'une variable aléatoire ou d'un système de variables aléatoires. Lorsque le nombre d'expériences est insuffisant on ne peut déterminer ces paramètres d'une manière plus ou moins précise; dans ces cas les données expérimentales sont dans une forte mesure aléatoires, donc tous les paramètres calculés à partir de ces données se trouvent également aléatoires. Dans ces conditions il peut seulement être question des « estimations » des paramètres cherchés, c'est-à-dire des valeurs approchées telles que, appliquées dans un grand nombre de cas, elles conduiraient en moyenne à des erreurs moindres que toutes autres. La recherche des valeurs approchées des caractéristiques numériques est étroitement liée à l'estimation de leur précision et de leur fiabilité. Nous rencontrerons des problèmes de ce genre dans le chapitre 14.

Nous avons là une liste, loin d'être complète, des problèmes fondamentaux des mathématiques statistiques. Nous n'avons énuméré que les problèmes les plus importants pour les applications pratiques. Dans le présent chapitre nous allons exposer quelques problèmes des plus élémentaires de la statistique et les méthodes de leur résolution.

7.2. Ensemble statistique simple. Fonction de répartition statistique

Soit une certaine variable aléatoire X, dont la loi de répartition est inconnue. Il y a lieu de trouver cette loi à partir de l'expérience ou de vérifier expérimentalement l'hypothèse suivant laquelle X suit telle ou telle loi. A cet effet la variable aléatoire X est soumise à une série d'expériences indépendantes (observations). Dans chacune de ces expériences la variable aléatoire X prend une certaine valeur. L'ensemble des valeurs observées de cette variable n'est autre chose que les données statistiques primaires devant être traitées, étudiées, analysées. Un tel ensemble est appelé « ensemble statistique simple » ou « suite statistique simple ». Il est d'usage de rassembler ces valeurs dans un tableau dans la première colonne duquel se trouve le numéro de l'expérience i et dans la seconde, la valeur observée de la variable aléatoire.

E x e m p l e 1. La variable aléatoire β est l'angle de glissement *) d'un avion lors du parachutage. Vingt parachutistes ont sauté, et chaque fois l'angle β a été enregistré en millièmes de radian. Les résultats des observations sont rassemblés dans le tableau.

i	β_i	i	β_i	i	β_i
1	−20	8	−30	15	−10
2	−60	9	120	16	20
3	−10	10	−100	17	30
4	30	11	−80	18	−80
5	60	12	20	19	60
6	70	13	40	20	70
7	−10	14	−60		

La suite statistique simple est une forme primaire d'écriture des données statistiques et peut être traitée de différentes manières. L'une des méthodes de traitement est la recherche de la *fonction de répartition statistique* d'une variable aléatoire.

On appelle fonction de répartition statistique (*ou empirique*) *de la variable aléatoire X la fréquence de l'événement* $X < x$ *dans les données statistiques*:

$$F^*(x) = P^*(X < x). \tag{7.2.1}$$

Pour trouver la valeur de la fonction de répartition statistique pour x donné il suffit de compter le nombre d'expériences dans lesquelles $X < x$ et de le diviser par le nombre total n d'expériences effectuées.

E x e m p l e 2. Construire la fonction de répartition statistique pour la variable aléatoire β de l'exemple 1.

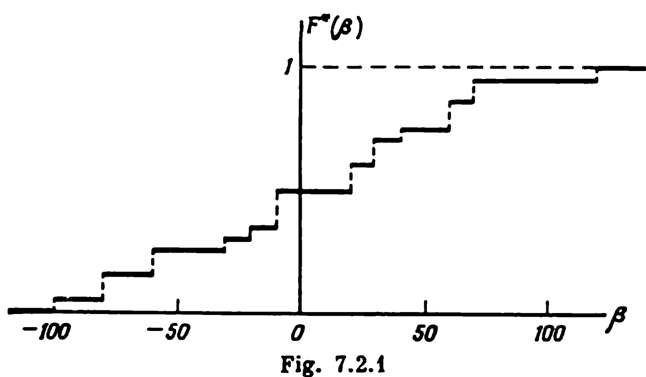


Fig. 7.2.1

S o l u t i o n. La valeur minimale observée de β étant −100, on a $F(-100) = 0$. La valeur −100 a été observée une fois, sa fréquence est donc

*) On entend par « angle de glissement » l'angle formé par le vecteur vitesse et le plan de symétrie de l'avion.

égale a 1/20 ; par conséquent au point —100 la fonction F^* (β) fait un saut égal à 1/20. Dans l'intervalle de —100 à —80 la fonction F^* (β) demeure égale à 1/20 ; au point —80 la fonction F^* (β) fait un saut de 2/20 car la valeur —80 a été observée deux fois, etc.

Le graphique de la fonction F^* (β) est donné sur la figure 7.2.1.

La fonction de répartition statistique d'une variable aléatoire quelconque, continue ou discrète, est toujours une fonction en escalier dont les échelons correspondent aux valeurs observées de la variable aléatoire et sont égaux en grandeur aux fréquences de ces valeurs. Si chaque valeur d'une variable aléatoire a été observée exactement une fois, le saut de la fonction de répartition statistique pour chaque valeur observée est égal à $1/n$, où n est le nombre d'observations.

Lorsque le nombre d'expériences n augmente, conformément au théorème de Bernoulli, pour x quelconque la fréquence de l'événement $X < x$ tend (en probabilité) vers la probabilité de cet événement. Par conséquent, lorsque n augmente la fonction de répartition $F^*(x)$ tend (en probabilité) vers la vraie fonction de répartition $F(x)$ de la variable aléatoire X.

Si X est une variable aléatoire continue, lorsque le nombre d'observations n augmente, le nombre de sauts de la fonction $F^*(x)$ augmente également, la valeur de ces sauts diminue et le graphique de la fonction $F^*(x)$ tend indéfiniment vers la courbe continue $F(x)$ de la fonction de répartition de X.

En principe, trouver la fonction de répartition statistique c'est déjà dépouiller les données expérimentales. Cependant lorsque le nombre d'expériences n est important, il est difficile de trouver $F^*(x)$ par la méthode décrite ci-dessus. De plus, souvent pour plus de clarté il est commode d'utiliser d'autres caractéristiques des répartitions statistiques, analogues non pas à la fonction de répartition $F(x)$ mais à la densité $f(x)$. Nous verrons dans le paragraphe suivant des méthodes de ce genre de description des données statistiques.

7.3. Suite statistique. Histogramme

Lorsque le nombre d'observations est important (de l'ordre de centaines) un ensemble statistique simple n'est plus une forme commode de description des données statistiques, car il devient volumineux et peu clair. Pour présenter les données statistiques sous une forme plus compacte et ordonnée on les soumet à un traitement complémentaire, en l'occurrence on établit une suite statistique.

Supposons que nous ayons à notre disposition les résultats des observations sur la variable aléatoire continue X sous la forme d'un ensemble statistique simple. Nous allons diviser la gamme des valeurs observées de X en intervalles, ou rangs, et compter le nombre de valeurs m_i correspondant à chaque i-ème rang. Puis, divisant ce nombre par le nombre total d'observations n nous trouvons la fré-

quence correspondant à un rang donné:

$$p_i^* = \frac{m_i}{n}. \qquad (7.3.1)$$

La somme des fréquences de tous les rangs doit évidemment être égale à l'unité.

Il est d'usage de rassembler les fréquences dans un tableau où les rangs sont disposés dans l'ordre croissant. On appelle ce tableau *suite statistique*:

I_i	x_1 ; x_2	x_2; x_3	...	x_i ; x_{i+1}	...	x_k ; x_{k+1}
p_i^*	p_1^*	p_2^*	...	p_i^*	...	p_k^*

I_i désigne ici l'i-ème rang; x_i, x_{i+1} ses limites; p_i^* la fréquence correspondante; k le nombre de rangs.

Exemple 1. On effectue 500 mesures de l'erreur de pointage en dérive lors du tir à partir d'un avion sur une cible terrestre. Les résultats des mesures (en millièmes de radian) sont rassemblés dans le tableau:

I_i	−4 ; −3	−3 ; −2	−2 ; −1	−1 ; 0	0 ; 1	1 ; 2	2 ; 3	3 ; 4
m_i	6	25	72	133	120	88	46	10
p_i^*	0,012	0,050	0,144	0,266	0,240	0,176	0,092	0,020

I_i désigne ici l'intervalle des valeurs des erreurs de pointage; m_i le nombre d'observations dans l'intervalle; $p_i^* = \frac{m_i}{n}$ la fréquence correspondante.

Lors du groupement des valeurs observées de la variable aléatoire il faut souvent décider où rapporter une valeur se trouvant à la limite de deux rangs. Dans de tels cas il est conseillé de considérer la valeur en question comme appartenant en égale mesure aux deux rangs et d'ajouter aux nombres m_i des deux rangs la quantité 1/2.

Le nombre de rangs de groupement des données statistiques ne doit pas être trop important (autrement la suite devient peu représentative et les fréquences accusent une irrégularité) ; il ne doit pas non plus être trop petit (lorsque le nombre de rangs est trop petit les propriétés de la répartition ne sont décrites que trop grossièrement par la suite statistique). L'expérience montre que dans la majorité des cas le nombre de rangs doit se trouver entre 10 et 20. Plus les données statistiques sont complètes et homogènes, plus grand peut

être le nombre de rangs de la suite statistique. Les rangs peuvent être de même longueur ou de longueurs différentes. Il est évidemment plus simple de les prendre de même longueur. Cependant lors du traitement des données sur les variables aléatoires réparties d'une manière très irrégulière, il est parfois commode de prendre dans le domaine de densité élevée des intervalles plus etroits que dans le domaine de densité faible.

Les suites statistiques sont souvent représentées graphiquement sous la forme d'*histogrammes*. Pour construire un histogramme on porte sur l'axe des abscisses les intervalles de classement sur chacun desquels, pris pour base, on construit un rectangle d'aire égale à la

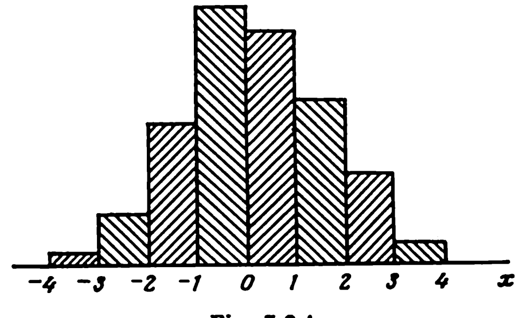


Fig. 7.3.1

fréquence du rang en question. Evidemment, la hauteur de chaque rectangle se trouve par division de la fréquence du rang par sa longueur. Lorsque les rangs sont de même longueur, les hauteurs des rectangles sont proportionnelles aux fréquences correspondantes. Il découle de la méthode de construction de l'histogramme que son aire doit être égale à l'unité.

A titre d'exemple on peut construire l'histogramme à partir des données de l'exemple 1 (fig. 7.3.1).

Il est évident qu'au fur et à mesure de l'augmentation du nombre d'expériences on peut prendre les intervalles de classement de plus en plus petits; l'histogramme se rapprochera alors de plus en plus d'une certaine courbe limitant une surface d'aire égale à l'unité. Il est facile de voir que cette courbe n'est rien d'autre que le graphique de la densité de probabilité de la variable X.

En utilisant les données de la suite statistique on peut également trouver approximativement la fonction de répartition statistique de X. Il serait trop coûteux et inutile de chercher la fonction de répartition exacte avec ses plusieurs centaines de paliers pour toutes les valeurs observées de X. Dans des fins pratiques il suffit de construire la fonction de répartition par plusieurs points. Ces points peuvent être les limites x_1, x_2, . . . des intervalles figurant dans la suite sta-

tistique. On a alors de toute évidence:

$$\left.\begin{aligned} &F^*(x_1)=0;\\ &F^*(x_2)=p_1^*;\\ &F^*(x_3)=p_1^*+p_2^*;\\ &\dots\dots\dots\dots\\ &F^*(x_k)=\sum_{i=1}^{k-1}p_i^*;\\ &F^*(x_{k+1})=\sum_{i=1}^{k}p_i^*=1. \end{aligned}\right\} \qquad (7.3.2)$$

En réunissant les points par une ligne brisée ou par une courbe continue on obtient le graphique approché de la fonction de répartition.

Exemple 2. Construire la fonction de répartition statistique approchée de l'erreur de pointage d'après les données de l'exemple 1.

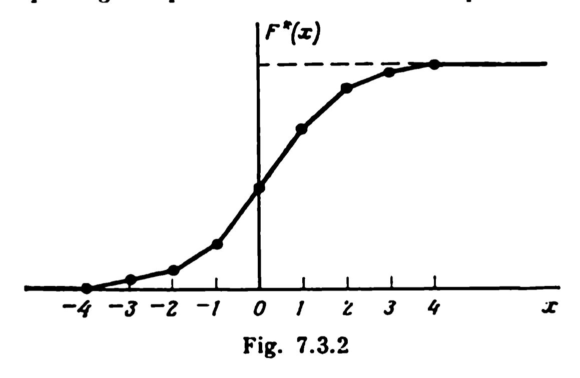


Fig. 7.3.2

Solution. D'après les formules (7.3.2) on a:

$F^*(-4)=0$; $F^*(-3)=0{,}012$; $F^*(-2)=0{,}012+0{,}050=0{,}062$;

$F^*(-1)=0{,}206$; $F^*(0)=0{,}472$; $F^*(1)=0{,}712$;

$F^*(2)=0{,}888$;

$F^*(3)=0{,}980$; $F^*(4)=1{,}000$.

Sur la figure 7.3.2 on peut voir le graphique de la fonction de répartition statistique approchée.

7.4. Caractéristiques numériques d'une répartition statistique

Au chapitre 5 nous avons étudié différentes caractéristiques numériques des variables aléatoires, à savoir l'espérance mathématique, la variance, les moments initiaux et centrés d'ordre différents. Ces carac-

téristiques numériques jouent un rôle important dans la théorie des probabilités. A toute caractéristique numérique de la variable aléatoire X correspond une caractéristique analogue statistique. A la caractéristique fondamentale de position, l'espérance mathématique d'une variable aléatoire, correspond la moyenne arithmétique des valeurs observées:

$$M^*[X]=\frac{\sum_{i=1}^{n} x_i}{n}, \tag{7.4.1}$$

où x_i est la valeur observée dans l'i-ème expérience, n le nombre d'expériences.

Dans la suite nous allons appeler cette caractéristique *moyenne statistique* d'une variable aléatoire.

En vertu de la loi des grands nombres, lorsque le nombre d'expériences augmente la moyenne statistique tend (en probabilité) vers l'espérance mathématique. Pour un n suffisamment grand la moyenne statistique peut approximativement être prise égale à l'espérance mathématique. Lorsque le nombre d'expériences est peu élevé, la moyenne statistique est une variable aléatoire qui néanmoins est liée à l'espérance mathématique et permet de s'en faire une assez bonne idée.

Des caractéristiques statistiques analogues existent pour toutes caractéristiques numériques. Nous les marquons d'un astérisque pour les distinguer des caractéristiques numériques correspondantes.

Soit par exemple la variance d'une variable aléatoire, c'est l'espérance mathématique de la variable aléatoire $\overset{\circ}{X}{}^2=(X-m_x)^2$:

$$D[X]=M[\overset{\circ}{X}{}^2]=M[(X-m_x)^2]. \tag{7.4.2}$$

Si dans cette expression on remplace l'espérance mathématique par la moyenne arithmétique on obtient la variance statistique de X:

$$D^*[X]=\frac{\sum_{i=1}^{n}(x_i-m_x^*)^2}{n}, \tag{7.4.3}$$

où $m_x^*=M^*[X]$ est la moyenne statistique.

D'une manière analogue on définit les moments statistiques initiaux et centrés d'ordre quelconque:

$$\alpha_s^*[X]=\frac{\sum_{i=1}^{n} x_i^s}{n}, \tag{7.4.4}$$

$$\mu_s^*[X]=\frac{\sum_{i=1}^{n}(x_i-m_x^*)^s}{n} \tag{7.4.5}$$

Toutes ces définitions sont analogues à celles du chapitre 5 pour les caractéristiques numériques des variables aléatoires, avec cette seule différence qu'au lieu de l'espérance mathématique on a partout la moyenne arithmétique. Lorsque le nombre d'observations augmente, il est clair que toutes les caractéristiques statistiques convergeront en probabilité vers les caractéristiques mathématiques correspondantes et pour n suffisamment grand peuvent être confondues avec celles-ci.

On peut facilement montrer que les moments initiaux et centrés statistiques possèdent les mêmes propriétés que les moments mathématiques. En particulier, le moment centré statistique du premier ordre est toujours nul:

$$\mu_1^* = \frac{\sum_{i=1}^{n} (x_i - m_x^*)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - m_x^* = m_x^* - m_x^* = 0.$$

Les relations signalées entre les moments centrés et initiaux restent également les mêmes:

$$\mu_2^* = D_x^* = \frac{\sum_{i=1}^{n} (x_i - m_x^*)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - 2m_x^* \frac{\sum_{i=1}^{n} x_i}{n} + (m_x^*)^2 = \alpha_2^* - (m_x^*)^2, \tag{7.4.6}$$

etc.

Lorsque le nombre d'expériences est très important le calcul des caractéristiques d'après les formules (7.4.1)-(7.4.5) devient trop compliqué. On peut alors recourir à une méthode approchée qui consiste à répartir les données par intervalles de classement, tout comme ceci a été fait pour la construction de la suite statistique ou de l'histogramme, et à considérer qu'à l'intérieur de chaque intervalle la valeur de la variable aléatoire est constante et égale à sa valeur moyenne qui devient ainsi la valeur représentative ou le « représentant » de l'intervalle. Les caractéristiques statistiques numériques s'exprimeront alors approximativement par les formules:

$$m_x^* = M^*[X] = \sum_{i=1}^{k} \tilde{x}_i p_i^*, \tag{7.4.7}$$

$$D_x^* = D^*[X] = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^2 p_i^*, \tag{7.4.8}$$

$$\alpha_s^*[X] = \sum_{i=1}^{k} \tilde{x}_i^s p_i^*, \tag{7.4.9}$$

$$\mu_s^*[X] = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^s p_i^*, \tag{7.4.10}$$

ici $\tilde{x}_i$ est la valeur représentative du i-ème intervalle; p_i^* la fréquence du i-ème intervalle; k le nombre d'intervalles.

On voit que les formules (7.4.7)-(7.4.10) sont analogues aux formules des §§ 5.6 et 5.7 définissant l'espérance mathématique, la variance, les moments initiaux et centrés de la variable aléatoire continue X, avec cette seule différence que les probabilités p_i y sont remplacées par les fréquences p_i^*, l'espérance mathématique m_x par la moyenne statistique m_x^*, le nombre de valeurs possibles de la variable aléatoire par le nombre d'intervalles.

La majorité des cours de théorie des probabilités et de statistique mathématique utilisent pour les caractéristiques statistiques analogues aux caractéristiques numériques des variables aléatoires une terminologie différente; plus exactement la moyenne statistique est appelée « moyenne échantillonnée », la variance statistique « variance échantillonnée », etc. L'origine de ces termes peut s'expliquer comme suit: en statistique, surtout rurale et biologique, on utilise souvent des distributions de tel ou tel indice pour un ensemble important d'éléments (d'individus), formant une collectivité statistique (cet indice peut par exemple être la teneur en albumen des graines de blé, le poids de ces mêmes graines, la taille et le poids d'une bête, etc.). Cet indice est une variable aléatoire dont la valeur change d'un individu à l'autre. Cependant pour pouvoir se faire une idée de la répartition de cette variable aléatoire ou de ses caractéristiques principales il n'est pas indispensable d'étudier chaque individu de la collectivité ; on peut étudier un certain échantillon suffisamment grand pour qu'apparaissent les traits essentiels de la loi étudiée. L'ensemble dans lequel on prélève l'échantillon s'appelle en statistique *grand ensemble*. On suppose que le nombre d'éléments (d'individus) N de l'ensemble est très important, et le nombre d'éléments n d'un échantillon est limité. Lorsque N est suffisamment grand les propriétés des distributions échantillonnées (statistiques) et des caractéristiques ne dépendent pratiquement pas de N ; d'où l'idéalisation mathématique en vertu de laquelle l'ensemble dont sont issus les échantillons est de taille infinie. On distingue alors les caractéristiques exactes (lois de répartition, espérance mathématique, variance, etc.) du grand ensemble des caractéristiques analogues échantillonnées. Les caractéristiques échantillonnées diffèrent des caractéristiques correspondantes du grand ensemble par suite de la taille limitée n de l'échantillon; lorsque n augmente indéfiniment, il est évident que toutes les caractéristiques tendent (convergent en probabilité) vers les caractéristiques correspondantes du grand ensemble. Souvent il y a lieu de trouver la taille n de l'échantillon suffisante pour pouvoir déterminer avec une précision requise à partir des caractéristiques échantillonnées, les caractéristiques du grand ensemble ou bien la précision avec laquelle, pour une taille donnée de l'échantillon, on peut connaître les caractéristiques du grand en-

semble. Cette méthode d'une étude parallèle d'un grand ensemble infini, dans lequel on prélève un échantillon, et d'un échantillon de taille limitée est tout à fait naturelle dans le cas où les collectivités étudiées sont vraiment très populeuses. Cependant dans un grand nombre de problèmes, il ne s'agit pas de choisir un échantillon dans un grand ensemble d'éléments mais spécialement effectuer une expérience pour étudier telle ou telle variable aléatoire, sa loi ou ses caractéristiques numériques. Par exemple, pour l'étude des erreurs de guidage par une station radar d'un appareil cosmique, il peut s'avérer nécessaire d'effectuer des expériences spéciales ; il en est de même lorsque des expériences sont réalisées dans le but de voir si tel ou tel facteur est nuisible à la vie d'un être vivant, ou bien quand on effectue des lancements de fusées dans le but de trouver les caractéristiques de dispersion, etc. Dans ces conditions le nombre d'expériences est limité non pas par des difficultés d'enregistrement ou de traitement, mais par la complexité et le coût de chaque expérience. Il est évident que dans ce cas également on peut envisager les n expériences effectuées comme un échantillon issu d'un grand ensemble quelconque constitué d'un nombre infiniment grand d'expériences possibles, réalisables en principe dans les conditions en question. Cependant ce point de vue n'est pas justifié par la nature des problèmes étudiés et serait une idéalisation inutile. C'est pourquoi dans le présent ouvrage nous n'utilisons pas les termes traditionnels: « moyenne échantillonnée », « variance échantillonnée », « caractéristiques échantillonnées », etc. en employant à leur place les termes équivalents de « moyenne statistique », « variance statistique », « caractéristiques statistiques », etc.

7.5. Approximation des données statistiques

Dans toute distribution statistique le hasard intervient dans une mesure plus ou moins grande, ceci étant lié au fait que le nombre d'expériences est limité, que les expériences effectuées ont été choisies au hasard parmi d'autres et donc les résultats qu'elles ont fournis sont également aléatoires. Ce n'est que si le nombre d'expériences est très grand que le hasard joue un rôle de moins en moins important et dans le phénomène aléatoire se manifeste alors nettement la loi qui lui est propre. Dans la pratique on n'a presque jamais affaire à un nombre aussi important d'expériences et on est obligé à compter avec le hasard. C'est pourquoi lors du traitement des données statistiques il importe de savoir choisir la courbe théorique approchant une suite statistique de façon à exprimer les traits essentiels des données statistiques mais non les traits accidentels provenant d'une insuffisance de données expérimentales. C'est ce qui s'appelle l'*approximation des données statistiques*.

Le problème consiste à trouver une courbe théorique continue, qui

d'un certain point de vue donne une meilleure description des données statistiques en question (fig. 7.5.1).

La recherche de la meilleure approximation des données statistiques, tout comme n'importe quel problème de meilleure représentation analytique d'une fonction empirique, est un problème assez vaguement déterminé et sa solution dépend de ce qu'on entend par « meilleur ». Par exemple, très répandue est la méthode dite des moindres carrés (voir § 14.5) suivant laquelle la meilleure approximation d'une relation empirique pour une classe donnée de fonctions est

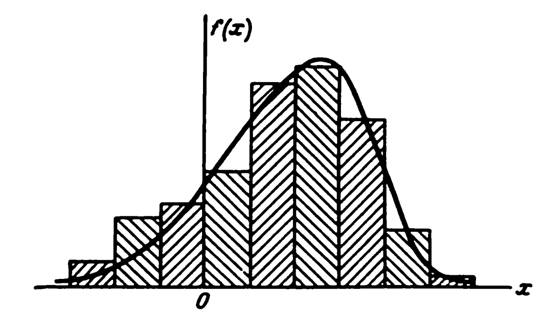


Fig. 7.5.1

celle où la somme des carrés des écarts est minimale. La classe de fonctions où l'on cherche cette meilleure approximation est choisie non pas à partir des considérations mathématiques mais en se basant sur le sens physique du problème à résoudre, compte tenu du caractère de la courbe empirique obtenue et de la précision des observations. Souvent on connaît d'avance la forme de la fonction exprimant la relation étudiée, l'expérience devant seulement fournir un certain nombre de paramètres numériques entrant dans l'expression de la fonction; ce sont justement ces paramètres que l'on détermine par la méthode des moindres carrés.

Il en est de même du problème d'approximation des données statistiques. En général, la forme de la courbe théorique est choisie d'avance à partir de considérations physiques et dans certains cas elle est prédéterminée par la répartition statistique. L'approximation des données statistiques revient au choix rationnel des valeurs des paramètres de la courbe approchée pour lesquels les distributions statistique et théorique se correspondent le mieux.

Soit par exemple X l'erreur de mesure provenant de l'action conjointe d'un grand nombre d'erreurs élémentaires; des considérations théoriques suggèrent que X suit une loi normale de densité

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}}, \qquad (7.5.1)$$

et le problème du lissage revient à celui du choix rationnel des paramètres m et σ dans l'expression (7.5.1).

Parfois on sait d'avance que X est répartie à peu près uniformément sur un certain intervalle; le problème consiste alors à choisir convenablement les paramètres de la loi de densité uniforme:

$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{pour} \quad \alpha < x < \beta, \\ 0 & \text{pour} \quad x < \alpha \text{ ou } x > \beta \end{cases}$$

pouvant approcher de la meilleure façon la répartition statistique en question.

Il ne faut pas oublier que toute fonction analytique $f(x)$ permettant d'approcher une distribution empirique doit avoir les propriétés essentielles d'une densité de probabilité, à savoir:

$$\left. \begin{aligned} & f(x) \geqslant 0; \\ & \int_{-\infty}^{\infty} f(x)\,dx = 1. \end{aligned} \right\} \qquad (7.5.2)$$

Soit $f(x)$ une fonction satisfaisant aux conditions (7.5.2), choisie pour approcher une certaine répartition statistique; cette fonction contient plusieurs paramètres a, b, . . .,; il y a lieu de trouver ces paramètres de telle sorte que la fonction $f(x)$ décrive au mieux les données statistiques en question. L'une des méthodes utilisées pour résoudre ce problème est la *méthode des moments*.

En vertu de la méthode des moments, les paramètres a, b, . . . sont choisis de telle sorte que plusieurs caractéristiques numériques importantes (moments) de la répartition théorique soient égales aux caractéristiques statistiques correspondantes. Par exemple, si la courbe théorique $f(x)$ comporte seulement deux paramètres a et b, on les choisit de façon à assurer que l'espérance mathématique m_x et la variance D_x de la répartition théorique soient égales aux caractéristiques statistiques m_x^* et D_x^* correspondantes. Si la courbe $f(x)$ comporte trois paramètres on peut les choisir de telle sorte que coïncident les trois premiers moments, etc. Très utile dans la description des données statistiques peut s'avérer un système spécial de *courbes de Pearson* dont chacune dépend dans le cas général de quatre paramètres, lesquels sont choisis de manière à conserver les quatre premiers moments de la répartition statistique (espérance mathématique, variance, moments trois et quatre). N. Borodatschev *) a donné un système fort intéressant de courbes de répartition en se basant sur un autre principe. Il a proposé de choisir le type de la courbe théo-

*) N. Borodatschev. *Fondements de la théorie de la précision industrielle* (en russe). Ed. de l'Acad. des Sciences de l'U.R.S.S., Moscou, 1950.

rique non pas à partir des indices formels, mais en analysant la nature physique du phénomène ou du processus aléatoire étudié.

Il faut noter que dans la description des données statistiques l'utilisation des moments d'ordre supérieur au quatrième est dénuée de sens, car la précision du calcul des moments diminue rapidement lorsque leur ordre augmente.

Exemple 1. Dans l'exemple 1 du § 7.3 nous avons donné la répartition statistique de l'erreur de pointage en dérive X lorsqu'un avion tire sur une cible terrestre. Approcher cette répartition par la loi normale:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Solution. La loi normale dépend de deux paramètres: m et σ. Choisissons ces paramètres de façon à conserver les deux premiers moments — l'espérance mathématique et la variance — de la répartition statistique.

Calculons approximativement la moyenne statistique de l'erreur de pointage par la formule (7.4.7), en prenant pour valeur représentative de chaque intervalle sa valeur de milieu, il vient:

$$m_x^* = -3{,}5\cdot 0{,}012 - 2{,}5\cdot 0{,}050 - 1{,}5\cdot 0{,}144 - 0{,}5\cdot 0{,}266 + \\ + 0{,}5\cdot 0{,}240 + 1{,}5\cdot 0{,}176 + 2{,}5\cdot 0{,}092 + 3{,}5\cdot 0{,}020 = 0{,}168.$$

Pour trouver la variance nous allons calculer tout d'abord le moment initial deux par la formule (7.4.9) en posant $s = 2$, $k = 8$:

$$\alpha_2^* = \sum_{i=1}^{\infty} \tilde{x}_i^2 p_i^* = 2{,}126.$$

En exprimant la variance en fonction du moment initial deux d'après la formule (7.4.6) on a:

$$D_x^* = \alpha_2^* - (m_x^*)^2 = 2{,}126 - 0{,}028 = 2{,}098.$$

Choisissons les paramètres m et σ de la loi normale de telle sorte que soient remplies les conditions:

$$m = m_x^*; \quad \sigma^2 = D_x^*,$$

donc

$$m = 0{,}168; \quad \sigma = 1{,}448.$$

La loi normale s'écrit donc:

$$f(x) = \frac{1}{1{,}448\sqrt{2\pi}} e^{-\frac{(x-0{,}168)^2}{2\cdot 1{,}448^2}}$$

En utilisant la table 3 de l'annexe calculons les valeurs de $f(x)$ aux limites des intervalles:

x	-4	-3	-2	-1	0	1	2	3	4
$f(x)$	0,004	0,025	0,090	0,199	0,274	0,234	0,124	0,041	0,008

La figure 7.5.2 montre l'histogramme et la courbe de répartition approchée. On remarque que la courbe théorique $f(x)$, tout en conservant les particularités essentielles de la répartition statistique, est parfaitement régulière.

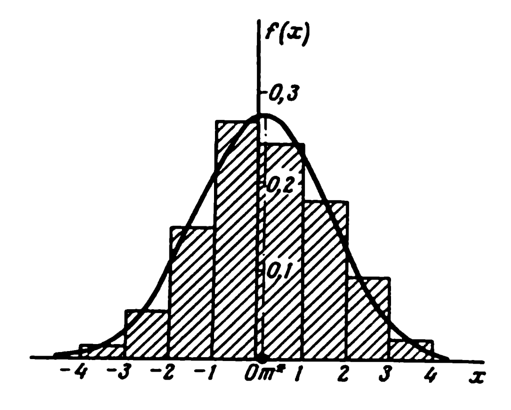


Fig. 7.5.2

Les irrégularités qu'accuse l'histogramme sont évidemment dues à des causes aléatoires; dans le paragraphe suivant nous allons donner un fondement plus sérieux à cette assertion.

N o t a : Dans le présent exemple lors du calcul de D_x^* nous avons utilisé l'expression (7.4.6) donnant la variance statistique en fonction du moment initial deux. On peut recommander cette méthode seulement si l'espérance mathématique m_x^* de la variable aléatoire étudiée X est peu importante; dans le cas contraire la formule (7.4.6) exprime la variance D_x^* en tant que différence de nombres voisins et se trouve être d'une précision médiocre. Dans ce cas on peut recommander soit de calculer D_x^* directement par la formule (7.4.3), soit de transférer l'origine des coordonnées en un point quelconque voisin de m_x^* et d'appliquer ensuite la formule (7.4.6). Utiliser la formule (7.4.3) équivaut à situer l'origine des coordonnées au point m_x^*; mais ceci n'est pas toujours commode car l'expression de m_x^* peut être fractionnaire, or, la soustraction d'une fraction de chaque x_i complique considérablement les calculs; dans ce cas il y a lieu de transporter l'origine des coordonnées en un point x d'abscisse entière, voisin de m_x^*.

E x e m p l e 2. Dans le but d'étudier la loi de répartition de l'erreur de mesure de la distance à l'aide d'un radioaltimètre on effectue 400 mesures de distance. Les résultats des expériences sont rassemblés dans le tableau :

I_i (m)	20 ; 30	30 ; 40	40 ; 50	50 ; 60	60 ; 70	70 ; 80	80 ; 90	90 ; 100
m_i	21	72	66	38	51	56	64	32
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080 .

Décrire ces données statistiques par la loi de densité uniforme.

S o l u t i o n. La loi de densité uniforme s'exprime par la formule:

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{pour} \quad \alpha<x<\beta,\\ 0 & \text{pour} \quad x<\alpha \text{ ou } x>\beta\end{cases}$$

et dépend de deux paramètres α et β. Ces paramètres doivent être choisis de manière à conserver les deux premiers moments de la répartition statistique, à savoir l'espérance mathématique m_x^* et la variance D_x^*. Exprimons l'espérance mathématique et la variance d'une loi de densité uniforme comme indiqué dans l'exemple du § 5.8:

$$m_x=\frac{\alpha+\beta}{2},$$

$$D_x=\frac{(\beta-\alpha)^2}{12}.$$

Pour simplifier les calculs liés à la détermination des moments statistiques plaçons l'origine des coordonnées au point $x_0 = 60$ et prenons pour valeur représentative de chaque intervalle la valeur de milieu. Le tableau ci-dessus devient:

$\tilde{x}'_i$	−35	−25	−15	−5	5	15	25	35
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,80

où $\tilde{x}'_i$ est la valeur moyenne de l'erreur X' du radioaltimètre dans un intervalle après le changement d'origine.

La valeur approchée de la moyenne statistique de l'erreur X' est:

$$m_{x'}^*=\sum_{i=1}^{k}\tilde{x}'_i p_i^*=0{,}26.$$

Le moment statistique deux de X' vaut:

$$\alpha_2^*=\sum_{i=1}^{k}(\tilde{x}'_i)^2\,p_i^*=447{,}8,$$

d'où la variance statistique:

$$D_{x'}^*=\alpha_2^*-(m_{x'}^*)^2=447{,}7.$$

En passant à l'origine initiale on obtient une nouvelle moyenne statistique:

$$m_x^*=m_{x'}^*+60=60{,}26$$

et la même variance statistique

$$D_x^* = D_{x'}^* = 447{,}7.$$

Les paramètres de la loi de densité uniforme sont donnés par les équations

$$\frac{\alpha+\beta}{2} = 60{,}26\,; \quad \frac{(\beta-\alpha)^2}{12} = 447{,}7.$$

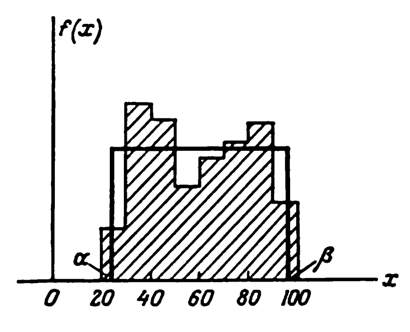


Fig. 7.5.3

La solution de ces équations par rapport à α et β donne

$$\alpha \approx 23{,}6\,; \quad \beta \approx 96.9,$$

d'où

$$f(x) = \frac{1}{\beta-\alpha} = \frac{1}{73{,}3} \approx 0{,}0136.$$

Sur la figure 7.5.3 on peut voir l'histogramme et la loi de densité uniforme $f(x)$ approchée.

7.6. Critères de conformité

Dans ce paragraphe nous allons étudier une des questions liées à la vraisemblance des hypothèses; il s'agit de la conformité des répartitions théorique et statistique.

Supposons qu'une répartition statistique soit approchée par la courbe théorique $f(x)$ (fig. 7.6.1). Même si la courbe théorique est bien choisie, certains écarts entre celle-ci et la répartition statistique sont inévitables. Une question se pose tout naturellement : ces écarts, sont-ils dus uniquement au hasard, vu le nombre limité d'observations, ou bien sont-ils essentiels et proviennent de ce que nous avons mal choisi la courbe de lissage de la répartition statistique étudiée ? Pour répondre à cette question on fait appel aux *critères de conformité*.

Voyons le sens à attribuer à la notion de critère de conformité. En se basant sur certaines données statistiques il y a lieu de vérifier l'hypothèse H selon laquelle la variable aléatoire X suit une certaine loi de répartition. Cette loi peut être donnée sous une forme quelconque : comme la fonction de répartition $F(x)$, comme la densité de

probabilité $f(x)$ ou comme un ensemble des probabilités p_i, où p_i est la probabilité pour X de tomber dans l'i-ème intervalle.

Etant donné que la forme $F(x)$ est la plus générale et détermine toutes les autres, nous allons considérer que l'hypothèse H sera justement en ce que la variable X a $F(X)$ pour fonction de répartition.

Pour adopter ou rejeter l'hypothèse H considérons une certaine grandeur U caractérisant la différence entre les répartitions statistique et théorique. La grandeur U peut être choisie de plusieurs façons;

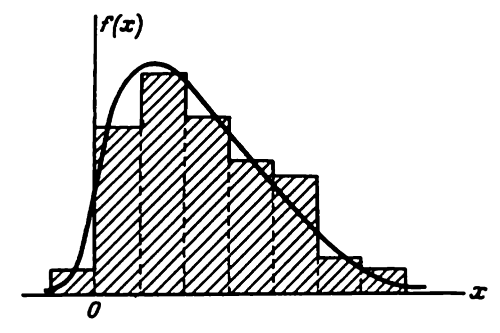


Fig. 7.6.1

par exemple, on peut prendre la somme des carrés des écarts des probabilités théoriques p_i des fréquences correspondantes p_i^*, ou la somme des mêmes carrés mais pondérés par certains coefficients (poids) ou l'écart maximal de la fonction de répartition statistique $F^*(x)$ de la fonction théorique $F(x)$, etc. Supposons que d'une manière ou d'une autre on ait choisi la grandeur U. Il est évident que c'est une certaine *variable aléatoire*. Sa loi de répartition dépend de celle de la variable aléatoire X ayant fait l'objet des expériences et du nombre d'expériences n. Si l'hypothèse H est vraie, la loi de répartition de U est déterminée par la loi de répartition de X (fonction $F(x)$) et le nombre n.

Supposons que nous connaissions cette loi de répartition. On a trouvé que dans une série d'expériences la mesure U que nous avons choisie pour évaluer l'écart a pris une certaine valeur u. La question qui se pose est de savoir si ceci peut s'expliquer par le hasard ou si cette valeur est trop importante et indique qu'il y a une différence essentielle entre les répartitions théorique et statistique et par conséquent l'hypothèse H n'est pas vraie. Pour répondre à cette question on suppose que l'hypothèse H est vraie et on calcule dans cette hypothèse la probabilite de l'événement suivant dû aux causes aléatoires:

$$U \geqslant u.$$

Si cette probabilité est petite on doit rejeter l'hypothèse H comme peu vraisemblable; si cette différence est importante, il y a lieu

d'adopter que les données expérimentales ne sont pas en contradiction avec l'hypothèse H.

On se pose alors la question de savoir comment il faut choisir la mesure U. Il se trouve que dans certains cas la loi de répartition de U a des propriétés fort simples et pour n suffisamment grand ne dépend pratiquement pas de la fonction $F(x)$. Ce sont ces mesures qui sont utilisées en mathématiques statistiques en qualité de critères de conformité.

Soit par exemple l'un des critères des plus fréquemment utilisés, à savoir le « critère du χ^2 » de Pearson.

Supposons que l'on ait effectué n expériences indépendantes dans chacune desquelles la variable aléatoire X a pris une certaine valeur. Les résultats des expériences sont classés en k intervalles et présentés sous la forme du tableau:

I_i	$x_1; x_2$	$x_2; x_3$	...	$x_k; k_{x+1}$
p_i^*	p_1^*	p_2^*	...	p_k^*

Il y a lieu de vérifier si les données expérimentales confirment l'hypothèse selon laquelle X suit la loi de répartition en question [donnée par la fonction de répartition $F(x)$ ou la densité $f(x)$]. Nous allons appeler cette loi de répartition « loi théorique ».

Connaissant la loi de répartition théorique on peut trouver les probabilités théoriques $p_1, p_2, \ldots, p_k$ pour la variable aléatoire étudiée de se trouver dans l'intervalle correspondant.

En vérifiant la conformité des répartitions théorique et statistique nous allons nous baser sur la différence entre les probabilités théoriques p_i et les fréquences observées p_i^*. Il est naturel de choisir pour mesure de la non-conformité des répartitions théorique et statistique la somme des carrés des écarts $(p_i^* - p_i)$ pris avec des « poids » c_i:

$$U = \sum_{i=1}^{k} c_i (p_i^* - p_i)^2. \tag{7.6.1}$$

Les coefficients c_i (« poids » des intervalles) tiennent compte de l'importance relative différente des écarts mentionnés dans divers intervalles. En effet un écart $p_i^* - p_i$ de même valeur absolue peut être peu important si la probabilité est grande, et notable si elle est petite. C'est pourquoi il serait naturel de prendre les « poids » c_i inversement proportionnels aux probabilités p_i des intervalles.

Alors il faut savoir choisir le coefficient de proportionnalité.

Pearson a montré que si l'on pose

$$c_i = \frac{n}{p_i}, \tag{7.6.2}$$

pour des n grands, la loi de répartition de U est douée de propriétés intéressantes: elle ne dépend pratiquement pas de la fonction de

répartition $F(x)$ ni du nombre d'expériences n, mais seulement du nombre d'intervalles de classement k en tendant vers la loi dite « du χ^2 » *).

Lorsque les coefficients c_i sont ainsi choisis, la mesure de la non-conformité est généralement désignée par χ^2:

$$\chi^2 = n \sum_{i=1}^{k} \frac{(p_i^* - p_i)^2}{p_i}. \tag{7.6.3}$$

Pour rendre les calculs plus commodes (pour ne pas avoir à manipuler avec des fractions avec un grand nombre de zéros) on peut introduire n sous le signe de la somme et, compte tenu de $p_i^* = \frac{m_i}{n}$, où m_i est le nombre de valeurs dans l'i-ème intervalle, écrire la formule (7.6.3) comme suit:

$$U = \chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}. \tag{7.6.4}$$

La loi du χ^2 dépend du paramètre r appelé *degré de liberté* de la répartition. Le nombre de degrés de liberté r est égal au nombre k de rangs moins celui des conditions indépendantes (contraintes) imposées aux fréquences p_i^*. Donnons quelques exemples de telles conditions:

$$\sum_{i=1}^{k} p_i^* = 1$$

si l'on exige seulement que la somme des fréquences relatives soit égale à l'unité (cette condition toujours imposée); ou

$$\sum_{i=1}^{k} \tilde{x}_i p_i^* = m_x$$

*) On appelle loi du χ^2 à r degrés de liberté la répartition des carrés de r variables aléatoires indépendantes, dont chacune est répartie suivant une loi normale d'espérance mathématique nulle et de variance unité La densité de cette répartition est

$$k_r(u) = \begin{cases} \dfrac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} & \text{pour } u > 0, \\ 0 & \text{pour } u < 0, \end{cases}$$

où $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$ est la fonction Γ.

si l'on choisit la répartition théorique de telle sorte que les moyennes des répartitions théorique et statistique coïncident; ou encore

$$\sum_{i=1}^{k}(\tilde{x}_i - m_x^*)^2 p_i^* = D_x$$

si l'on veut de plus que les variances théorique et statistique coïncident, etc.

Des tables spéciales ont éte établies pour la loi du χ^2 (voir table 4 de l'annexe). Ces tables permettent pour une valeur quelconque du χ^2 et un nombre r de degrés de liberté de trouver la probabilité p pour une variable répartie suivant la loi du χ^2 de dépasser cette valeur. Dans la table 4 on a porté comme paramètres la valeur de la probabilité p et le nombre r de degrés de liberté. Les cases de la table contiennent les valeurs correspondantes du χ^2.

La loi du χ^2 permet d'estimer le degré de conformité des répartitions théorique et statistique. Soit la variable X réellement répartie suivant la loi $F(x)$. La probabilité p trouvée dans la table est la probabilité pour que la mesure de la non-conformité (7.6.4) des répartitions théorique et statistique à cause des actions accidentelles soit inférieure à la valeur du χ^2 observée dans la série d'expériences en question. Si cette probabilité p est très petite (tellement petite que l'événement auquel elle se rapporte peut pratiquement être considéré comme impossible), il y a lieu d'admettre que le résultat de l'expérience est en contradiction avec l'hypothèse H selon laquelle $F(x)$ est la loi de répartition de X. Cette hypothèse doit être rejetée comme non vraisemblable. Au contraire, si la probabilité p est relativement importante, les écarts entre les répartitions théorique et statistique sont insignifiants et on peut les attribuer au hasard. L'hypothèse H selon laquelle X est répartie suivant la loi $F(x)$ doit être considérée vraisemblable ou tout au moins non contradictoire aux données expérimentales.

Ainsi l'utilisation du critère du χ^2 pour estimer la conformité des répartitions théorique et statistique se réduit à:

1) calculer la mesure de non-conformité χ^2 par la formule (7.6.4);

2) déterminer le nombre de degrés de liberté r en tant que nombre d'intervalles k moins le nombre de contraintes s:

$$r = k - s;$$

3) trouver dans la table 4 à partir de r et χ^2 la probabilité pour une variable répartie suivant la loi du χ^2 à r degrés de liberté d'être supérieure à une valeur donnée du χ^2. Si cette probabilité est assez petite l'hypothèse est rejetée comme non vraisemblable Si cette probabilité est relativement grande on peut adopter que l'hypothèse ne contredit pas les données expérimentales.

Il est difficile de dire combien la probabilité p doit être petite pour qu'il y ait lieu de rejeter et de modifier l'hypothèse; les raison-

nements mathématiques ne permettent pas de répondre à cette question, tout comme il est difficile de dire quelle doit être la valeur de la probabilité d'un événement pour que celui-ci soit pratiquement impossible. Dans la pratique, si p est inférieure à 0,1 il y a lieu de vérifier l'expérience, de la refaire si possible et si les écarts importants persistent, de rechercher une loi de répartition plus convenable pour décrire les données expérimentales en question.

Il faut remarquer que rares sont les cas où le critère du χ^2 (ou tout autre critère de conformité) permet de réfuter l'hypothèse avancée H et de la rejeter comme non conforme aux données expérimentales; si au contraire la probabilité p est grande, ceci ne peut en aucun cas être considéré comme une preuve irréfutable de la vraisemblance de l'hypothèse H, mais montre seulement que l'hypothèse ne contredit pas les données expérimentales.

A première vue on pourrait croire que plus la probabilité p est grande, meilleure est la conformité des répartitions théorique et statistique et plus fondé est le choix de la fonction $F(x)$ en tant que loi de répartition de X. En fait il n'en est rien. Supposons par exemple qu'en estimant la conformité des répartitions théorique et statistique par le critère du χ^2 on ait obtenu $p = 0{,}99$. Ceci signifie qu'avec une probabilité égale à 0,99, pour le nombre d'expériences effectuées, les écarts dus au hasard devraient être plus importants que ceux qui ont été observés. Mais les expériences ont fourni des écarts peu importants, trop petits pour que l'on puisse les considérer comme vraisemblables. Il semble plus raisonnable d'adopter qu'une coïncidence aussi bonne des répartitions théorique et statistique n'est pas due au hasard mais s'explique par l'arbitraire dans l'enregistrement et le traitement des données expérimentales (par exemple, par une « correction » des données expérimentales quand certains résultats sont rejetés à tort ou modifiés arbitrairement).

Il est évident que ces raisonnements ne sont vrais que si le nombre d'expériences n est suffisamment important (de l'ordre de plusieurs centaines) et lorsqu'il est raisonnable d'appliquer un critère basé sur la répartition limite de la mesure de non-conformité pour $n \to \infty$. Notons que pour pouvoir utiliser le critère du χ^2 il faut que non seulement le nombre d'expériences n soit important, mais également le nombre d'observations m_i dans chaque intervalle. Pratiquement il est conseillé d'avoir au moins 5-10 observations par intervalle. Si dans certains intervalles le nombre d'observations est assez petit (1 ou 2) il y a lieu de réunir plusieurs intervalles.

Exemple 1. Vérifier la conformité des répartitions théorique et statistique pour l'exemple 1 du § 7.5 (pages 129, 138).

Solution. En utilisant la loi de répartition théorique normale de paramètres

$$m = 0{,}168; \quad \sigma = 1\ 448$$

on trouve les probabilités de tomber dans chaque i-ème intervalle à l'aide de la formule:

$$p_i = \Phi^*\left(\frac{x_{i+1}-m}{\sigma}\right) - \Phi^*\left(\frac{x_i-m}{\sigma}\right),$$

où x_i, x_{i+1} sont les limites de l'i-ème intervalle.

On dresse ensuite le tableau comparatif des nombres m_i et des valeurs correspondantes np_i ($n = 500$):

I_i	−4; −3	−3; −2	−2; −1	−1; 0	0; 1	1; 2	2; 3	3; 4
m_i	6	25	72	133	120	88	46	10
np_i	6,2	26,2	71,2	122,2	131,8	90,5	38,2	10,5

Puis on calcule la valeur de la mesure de non-conformité par la formule (7.6.4):

$$\chi^2 = \sum_{i=1}^{8} \frac{(m_i - np_i)^2}{np_i} = 3{,}94.$$

On trouve alors le nombre de degrés de liberté comme le nombre d'intervalles moins le nombre de contraintes s (ici $s = 3$):

$$r = 8 - 3 = 5.$$

Dans la table 4 de l'annexe on trouve pour $r = 5$:

$$\text{pour } \chi^2 = 3{,}00 \quad p = 0{,}70;$$
$$\text{pour } \chi^2 = 4{,}35 \quad p = 0{,}50.$$

Par conséquent la probabilité cherchée p pour $\chi^2 = 3{,}94$ est approximativement égale à 0,56. Cette probabilité n'est pas petite; on peut donc considérer vraisemblable l'hypothèse selon laquelle la variable X suit une loi normale.

E x e m p l e 2. Vérifier la conformité des répartitions théorique et statistique pour les conditions de l'exemple 2 du § 7.5 (page 139).

S o l u t i o n. On calcule les valeurs p_i comme les probabilités pour une variable aléatoire uniformément répartie sur le segment (23,6; 96,9) de se trouver dans les intervalles (20; 30), (30; 40), etc. On dresse alors le tableau comparatif des valeurs m_i et np_i ($n = 400$):

I_i	20; 30	30; 40	40; 50	50; 60	60; 70	70; 80	80; 90	90; 100
m_i	21	72	66	38	51	56	64	32
np_i	34,9	54,6	54,6	54,6	54,6	54,6	54,6	38,0

La formule (7.6.4) donne χ^2:

$$\chi^2 = \sum_{i=1}^{8} \frac{(m_i - np_i)^2}{np_i} \approx 20{,}9$$

Le nombre de degrés de liberté est:

$$r = 8 - 3 = 5.$$

La table 4 de l'annexe donne pour $\chi^2 = 20,5$ et $r = 5$:

$$p = 0,001.$$

Par conséquent, l'écart observé entre les répartitions théorique et statistique ne pourrait être attribué au hasard qu'avec une probabilité $p \approx 0,001$. Cette probabilité étant très petite, il y a lieu d'admettre que les données expérimentales ne confirment pas l'hypothèse selon laquelle X suit une loi de densité uniforme.

En plus du critère du χ^2, d'autres critères sont couramment utilisés. Nous allons nous arrêter sur le *critère de Kolmogorov.*

A. Kolmogorov a pris pour mesure de non-conformité des répartitions théorique et statistique la valeur maximale du module de la différence entre la fonction de répartition statistique $F^*(x)$ et la fonction de répartition théorique correspondante, soit:

$$D = \max |F^*(x) - F(x)|.$$

Cette mesure de non-conformité est particulièrement simple à calculer; de plus assez simple est sa fonction de répartition. A. Kolmogorov a montré que, quelle que soit la fonction de répartition $F(x)$ d'une variable aléatoire continue X, lorsque le nombre d'observations indépendantes n augmente, la probabilité de l'inégalité

$$D\sqrt{n} \geqslant \lambda$$

tend vers

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}. \qquad (7.6.5)$$

Les valeurs de la probabilité $P(\lambda)$ calculées par la formule (7.6.5) sont rassemblées dans le tableau 7.6.1.

4 *Tableau 7.6.1*

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0,0	1,000	0,7	0,711	1,4	0,040
0,1	1,000	0,8	0,544	1,5	0,022
0,2	1,000	0,9	0,393	1,6	0,012
0,3	1,000	1,0	0,270	1,7	0,006
0,4	0,997	1,1	0,178	1,8	0,003
0,5	0,964	1,2	0,112	1,9	0,002
0,6	0,864	1,3	0,068	2,0	0,001

Le critère de Kolmogorov doit être utilisé comme suit: on construit la fonction de répartition statistique $F^*(x)$ et la courbe de

répartition théorique $F(x)$ et l'on détermine la quantité D comme le maximum du module de la différence des deux courbes (fig. 7.6.2).

Puis on calcule la grandeur

$$\lambda = D\sqrt{n}$$

et dans le tableau 7.6.1 on trouve la probabilité $P(\lambda)$. C'est la probabilité pour que, si X est vraiment répartie suivant la loi $F(x)$, l'écart maximal entre $F^*(x)$ et $F(x)$ dû aux facteurs aléatoires soit non

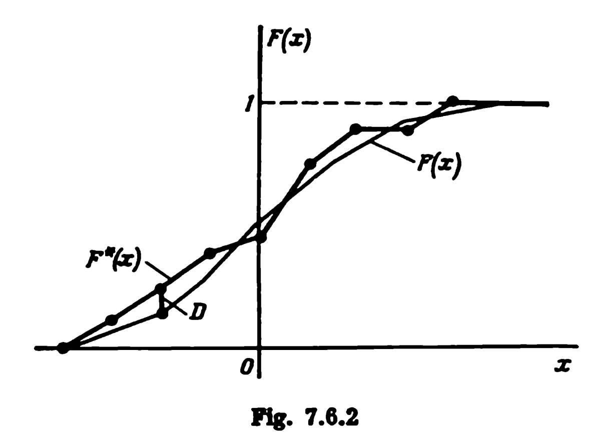


Fig. 7.6.2

inférieur à la valeur observée. Si la probabilité $P(\lambda)$ est petite, l'hypothèse doit être rejetée comme non vraisemblable; lorsque $P(\lambda)$ est assez grande on peut considérer qu'elle est compatible avec les données expérimentales.

Le critère de Kolmogorov a l'avantage d'être bien plus simple que celui du χ^2; c'est pour cela qu'il est souvent utilisé dans les applications pratiques. Il faut cependant noter que ce critère ne peut s'appliquer que si la répartition hypothétique $F(x)$ est donnée d'avance à partir de considérations théoriques, c'est-à-dire lorsqu'on connaît non seulement la forme de la fonction $F(x)$ mais également tous ses paramètres. C'est un cas assez rare dans la pratique. Généralement des considérations théoriques donnent seulement la forme de la fonction $F(x)$ alors que les paramètres numériques se calculent à partir des données statistiques. Pour le critère du χ^2 on en tient compte en diminuant le nombre de degrés de liberté de la répartition du χ^2. Le critère de Kolmogorov ne prévoit pas cette éventualité. Si l'on veut utiliser ce critère dans le cas où les paramètres de la répartition théorique sont calculés à partir des données statistiques, il donne des valeurs nettement exagérées de la probabilité $P(\lambda)$, de sorte que dans certains cas on risque d'adopter comme vraisemblable une hypothèse qui en réalité n est pas conforme aux données expérimentales.

CHAPITRE 8

SYSTÈMES DE VARIABLES ALÉATOIRES

8.1. Notion de système de variables aléatoires

Dans les applications pratiques de la théorie des probabilités on rencontre souvent des problèmes dans lesquels les résultats des expériences se trouvent décrits non pas par une mais par deux ou plus variables aléatoires formant un *système* ou un *ensemble*. Par exemple, le point d'impact d'un obus est déterminé non pas par une seule variable aléatoire, mais par deux : son abscisse et son ordonnée, lesquelles peuvent être considérées comme un système de deux variables aléatoires. De même le point d'éclatement d'un obus à distance est déterminé par un ensemble de trois variables aléatoires. Lors du tir d'ensemble de n coups les coordonnées des points d'impact sur un plan peuvent être considérées comme un ensemble de $2n$ variables aléatoires. L'éclatement d'un obus produit des éclats qui sont caractérisés par plusieurs variables aléatoires ; citons, par exemple le poids, les dimensions, la vitesse initiale, la direction de vol, etc. Nous désignerons par $(X, Y, \ldots, W)$ le système des variables aléatoires X. $Y, \ldots, W$.

Les propriétés d'un système de plusieurs variables aléatoires ne sont pas uniquement celles des composantes ; il faut de plus tenir compte de leur interdépendance.

Lors de l'étude des questions liées aux systèmes de variables aléatoires, il est commode d'utiliser l'interprétation géométrique des systèmes. Par exemple, un système de deux variables aléatoires (X, Y) peut être représenté par un point aléatoire dans le plan de coordonnées X et Y (fig. 8.1.1). D'une manière analogue un système de trois variables aléatoires peut être figuré par un point aléatoire dans un espace tridimensionnel. Plus généralement, on parle d'un système de n variables aléatoires comme d'un point aléatoire dans un espace à n dimensions. Bien que cette interprétation ne soit pas très figurative, elle est assez commode et permet de généraliser la terminologie et de simplifier les notations.

A côté de l'image du *point aléatoire* on utilise pour l'interprétation géométrique d'un système de variables aléatoires un *vecteur aléatoire*. Un système de deux variables aléatoires est considéré comme un vecteur aléatoire dans le plan xOy dont les composantes suivant

les axes sont des variables aléatoires X, Y (fig. 8.1.2). Un système de trois variables aléatoires est représenté par un vecteur aléatoire dans un espace tridimensionnel, un système de n variables aléatoires par un vecteur aléatoire dans un espace à n dimensions. La théorie des systèmes de variables aléatoires devient alors la *théorie des vecteurs aléatoires*.

Dans ce cours nous utiliserons l'une ou l'autre de ces interprétations suivant le problème étudié.

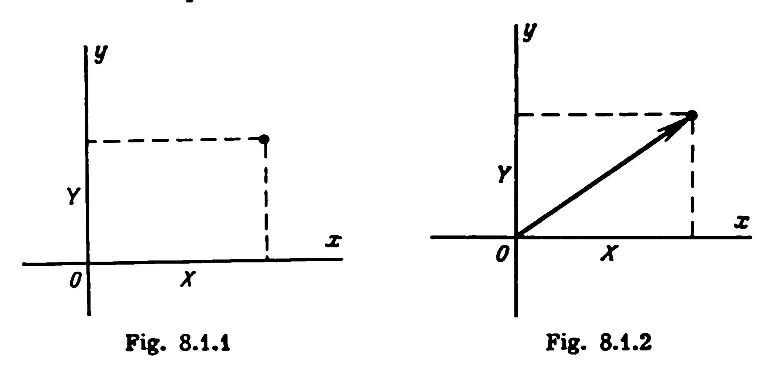


Fig. 8.1.1 Fig. 8.1.2

Lors de l'étude des systèmes de variables aléatoires nous envisagerons tant les caractéristiques probabilistes complètes, c'est-à-dire les lois de répartition, que les caractéristiques incomplètes, c'est-à-dire les caractéristiques numériques.

Nous allons commencer notre exposé par le cas simple d'un système de deux variables aléatoires.

8.2. Fonction de répartition d'un système de deux variables aléatoires

On appelle fonction de répartition d'un système de deux variabbles aléatoires (X, Y) la probabilité de vérification simultanée des deux inégalités $X < x$ et $Y < y$:

$$F(x, y) = P((X < x)(Y < y)). \tag{8.2.1}$$

En interprétant le système comme un point aléatoire, on voit que la fonction de répartition $F(x, y)$ n'est rien d'autre que la probabilité pour le point aléatoire (X, Y) de se trouver dans le quadrant infini de sommet au point (x, y), situé à gauche et en bas de celui-ci (fig. 8.2.1). Dans cette même interprétation la fonction de répartition de la variable aléatoire X seule, que nous désignerons par $F_1(x)$, est la probabilité pour le point aléatoire de se trouver dans le demi-plan limité à droite par l'abscisse x (fig. 8.2.2) ; de même la fonction de répartition $F_2(y)$ de Y est la probabilité pour le point aléatoire de

se trouver dans le demi-plan limité en haut par l'ordonnée y (fig. 8.2.3).

Dans le § 5.2 nous avons mentionné les propriétés essentielles de la fonction de répartition $F(x)$ d'une variable aléatoire unique. Nous allons formuler les propriétés analogues de la fonction de répartition

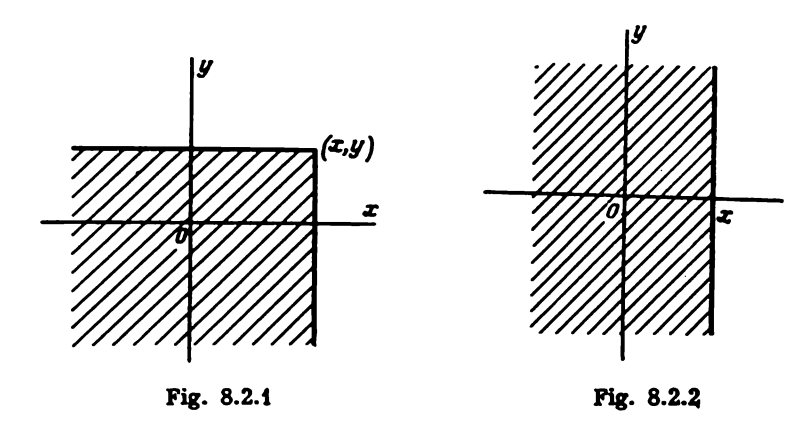


Fig. 8.2.1 Fig. 8.2.2

d'un système de variables aléatoires. A cet effet revenons à l'interprétation géométrique pour mieux illustrer ces propriétés.

1. La fonction de répartition $F(x, y)$ est une fonction non décroissante des deux arguments, c'est-à-dire

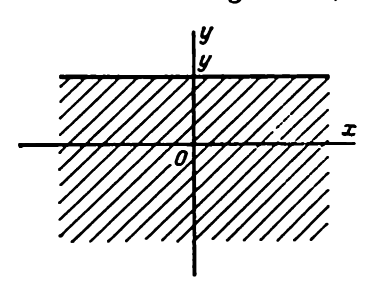


Fig. 8.2.3

pour $x_2 > x_1$ $F(x_2, y) \geqslant F(x_1, y)$;

pour $y_2 > y_1$ $F(x, y_2) \geqslant F(x, y_1)$.

L'interprétation géométrique de la fonction de répartition $F(x)$ en tant que probabilité pour le point (X, Y) de se trouver dans le quadrant de sommet (x, y) (fig. 8.2.1) permet facilement de s'en rendre compte. En effet, en augmentant x (par déplacement de la limite droite du quadrant à droite) ou en augmentant y (par déplacement de la limite supérieure vers le haut) on ne peut diminuer la probabilité de tomber dans ce quadrant.

2. Partout à $-\infty$ la fonction de répartition est égale à zéro:

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

Cette propriété apparaît lorsqu'on déplace indéfiniment à gauche la limite droite du quadrant ($x \to -\infty$) ou vers le bas sa limite supérieure ($y \to -\infty$), ou en effectuant cette opération simultanément

pour les deux limites; la probabilité de tomber dans le quadrant tend alors vers zéro.

3. Lorsque l'un des arguments est égal à $+\infty$, la fonction de répartition d'un système devient une fonction de répartition correspondant à l'autre argument:

$$F(x, +\infty) = F_1(x), \quad F(+\infty, y) = F_2(y),$$

où $F_1(x)$, $F_2(y)$ sont respectivement les fonctions de répartition des variables aléatoires X et Y.

On peut facilement s'en rendre compte en faisant $x \to +\infty$, ou $y \to +\infty$; à la limite le quadrant devient un demi-plan, la pro-

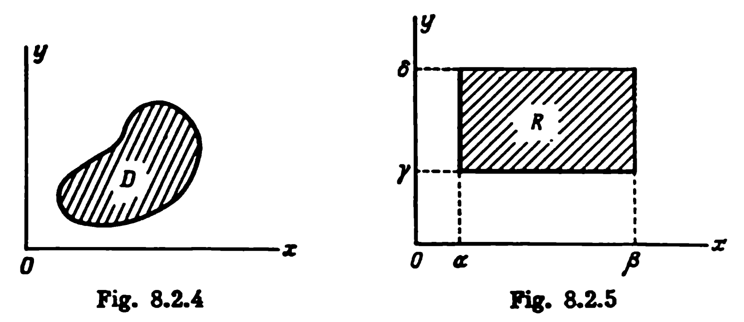


Fig. 8.2.4

Fig. 8.2.5

babilité de tomber dans ce demi-plan est donnée par la fonction de répartition de la variable respective.

4. Si les deux arguments sont égaux à $+\infty$, la fonction de répartition du système est égale à l'unité:

$$F(+\infty, +\infty) = 1.$$

En effet, pour $x \to +\infty$, $y \to +\infty$ le quadrant de sommet (x, y) devient à la limite le plan xOy, et le point s'y trouve certainement.

Lors de l'étude des lois de répartition des variables aléatoires indépendantes (chapitre 5) nous avons trouvé l'expression de la probabilité pour la variable aléatoire de se trouver dans l'intervalle donné. Nous avons exprimé cette probabilité tant à l'aide de la fonction de répartition qu'à l'aide de la densité de probabilité.

D'une manière analogue pour un système de deux variables aléatoires on peut trouver la probabilité pour le point aléatoire (X, Y) de se trouver dans un domaine donné D du plan xOy (fig. 8.2.4).

Convenons de désigner par le symbole $(X, Y) \subset D$ l'événement consistant en ce que le point aléatoire (X, Y) se trouve dans le domaine D.

La probabilité de cet événement s'exprime le plus simplement lorsque le domaine est un rectangle dont les côtés sont parallèles aux axes de coordonnées.

Exprimons à l'aide de la fonction de répartition du système la probabilité pour le point aléatoire (X, Y) de se trouver dans le rectangle R, limité par les abscisses α et β et les ordonnées γ et δ (fig. 8.2.5).

Il importe d'établir d'avance où rapporter les limites du rectangle. Tout comme nous l'avons fait dans le cas d'une variable aléatoire unique, nous considérons comme appartenantes au rectangle R sa limite inférieure et celle de gauche, sans y inclure ses limites supérieure et de droite *). Dans ce cas l'événement $(X, Y) \subset R$ est équivalent au produit des deux événements $\alpha \leqslant X < \beta$ et $\gamma \leqslant Y < \delta$. Exprimons la probabilité de cet événement à l'aide de la fonction de répartition du système. A cet effet considérons dans le plan xOy quatre quadrants infinis dont les sommets se trouvent aux points (β, δ) ; (α, δ) ; (β, γ) et (α, γ) (fig. 8.2.6).

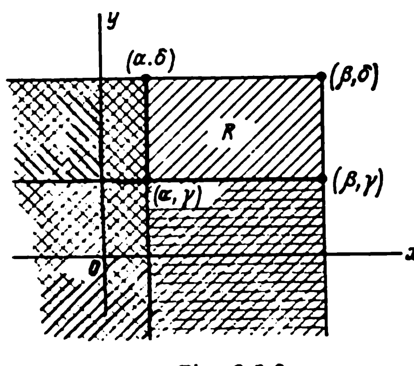


Fig. 8.2.6

La probabilité de tomber dans le rectangle R est de toute évidence égale à la probabilité de tomber dans le quadrant (β, δ) moins la probabilité de tomber dans le quadrant (α, δ) moins la probabilité de tomber dans le quadrant (β, γ) plus la probabilité de tomber dans le quadrant (α, γ) (car nous avons deux fois retranché la probabilité de tomber dans ce dernier quadrant). On trouve ainsi une formule donnant la probabilité de tomber dans le rectangle exprimée à l'aide de la fonction de répartition du système:

$$P((X, Y) \subset R) = F(\beta, \delta) - F(\alpha, \delta) - F(\beta, \gamma) + F(\alpha, \gamma). \tag{8.2.2}$$

Dans la suite, après avoir introduit la notion de densité de probabilité d'un système nous démontrerons la formule de la probabilité pour un point aléatoire de se trouver dans un domaine de forme arbitraire.

8.3. Densité de probabilité d'un système de deux variables aléatoires

La fonction de répartition, introduite dans le paragraphe précédent en tant que caractéristique complète d'un système existe pour des systèmes à variables aléatoires tant discontinues que continues.

*) Sur la figure 8.2.5 les limites incluses dans le rectangle sont données en trait gras.

Les systèmes de variables aléatoires continues jouent un rôle particulièrement important dans les applications pratiques. Les distributions des systèmes de variables aléatoires continues sont généralement caractérisées non pas par leur fonction de répartition mais par leur *densité de probabilité*.

Lorsque nous avons introduit la notion de densité de probabilité pour une seule variable aléatoire, nous l'avons définie comme la limite du rapport de la probabilité pour cette variable de se trouver dans un petit intervalle à la longueur de cet intervalle lorsque celui-ci diminue indéfiniment. D'une manière analogue on peut définir la densité de probabilité d'un système de deux variables aléatoires.

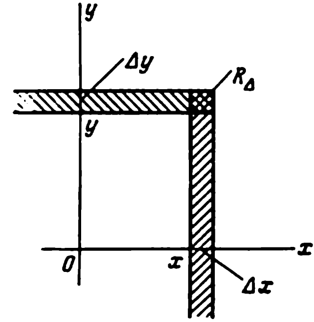


Fig. 8.3.1

Soit un système de deux variables aléatoires continues (X, Y) interprété comme un point aléatoire dans le plan xOy. Considérons dans ce plan un petit rectangle R_Δ dont les côtés sont Δx et Δy avec un sommet au point de coordonnées (x, y) (fig. 8.3.1). La probabilité de tomber dans ce rectangle en vertu de la formule (8.2.2) est

$$P((X, Y) \subset R_\Delta) = F(x + \Delta x, y + \Delta y) - F(x + \Delta x, y) - \\ - F(x, y + \Delta y) + F(x, y).$$

Divisons la probabilité de tomber dans le rectangle R_Δ par l'aire de ce rectangle et passons à la limite pour $\Delta x \to 0$ et $\Delta y \to 0$:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P((X, Y) \subset R_\Delta)}{\Delta x \Delta y} =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F(x+\Delta x, y+\Delta y) - F(x+\Delta x, y) - F(x, y+\Delta y) + F(x, y)}{\Delta x \, \Delta y}. \quad (8.3.1)$$

Supposons la fonction $F(x, y)$ non seulement continue mais également dérivable; le second membre de la formule (8.3.1) est alors la *dérivée partielle seconde mixte* de la fonction $F(x, y)$ par rapport à x et y. Désignons cette dérivée par $f(x, y)$:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \, \partial y} = F''_{xy}(x, y). \quad (8.3.2)$$

La fonction $f(x, y)$ est la *densité de probabilité* du système.

Ainsi la densité de probabilité d'un système est la limite du rapport de la probabilité de tomber dans un petit rectangle à l'aire de ce rectangle, lorsque ses deux dimensions tendent vers zéro; elle peut

s'exprimer comme la dérivée partielle seconde mixte de la fonction de répartition du système par rapport aux deux arguments.

Dans l'interprétation « mécanistique » de la distribution d'un système en tant que répartition d'une masse unité dans le plan xOy, la fonction $f(x, y)$ est la densité de probabilité de la masse au point (x, y).

Du point de vue géométrique la fonction $f(x, y)$ peut être une certaine surface (fig. 8.3.2). Cette surface est analogue à la courbe de répartition d'une variable aléatoire et s'appelle *surface de répartition*.

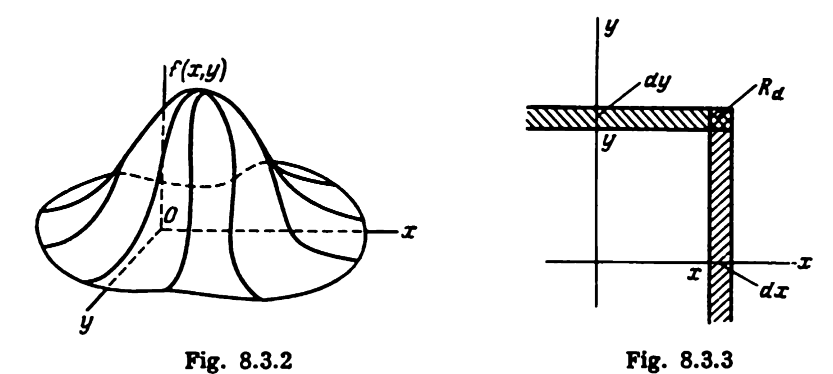


Fig. 8.3.2 Fig. 8.3.3

En coupant la surface de répartition $f(x, y)$ par un plan parallèle au plan xOy et en projetant la section obtenue sur le plan xOy on obtient une courbe dont la densité est constante en tous les points. Ces courbes sont appelées *courbes d'égale densité*. Les courbes d'égale densité sont de toute évidence des horizontales de la surface de répartition. Il est souvent commode de donner une répartition par une famille de courbes d'égale densité.

Lorsque nous avons étudié la densité de probabilité $f(x)$ d'une seule variable aléatoire nous avons introduit la notion d'« élément de probabilité » $f(x)\,dx$. C'est la probabilité pour la variable aléatoire X de se trouver dans l'intervalle élémentaire dx adjacent au point x. D'une manière analogue on peut introduire la notion d'« élément de probabilité » pour un système de variables aléatoires. Dans ce cas l'élément de probabilité est donné par l'expression;

$$f(x, y)\,dx\,dy.$$

Il est évident que l'élément de probabilité n'est rien d'autre que la probabilité de tomber dans le rectangle élémentaire de côtés dx, dy adjacent au point (x, y) (fig. 8.3.3).

Cette probabilité se mesure par le volume du parallélépipède élémentaire ayant pour base le rectangle élémentaire $dx\,dy$ et limité supérieurement par la surface $f(x, y)$ (fig. 8.3.4).

En utilisant la notion d'élément de probabilité on peut trouver l'expression de la probabilité pour un point aléatoire de se trouver dans un domaine quelconque D. Pour obtenir cette probabilité il y a évidemment lieu de prendre la somme (l'intégrale) des éléments de probabilité sur tout le domaine D:

$$P((X, Y) \subset D) = \iint_{(D)} f(x, y)\, dx\, dy. \qquad (8.3.3)$$

Du point de vue géométrique, la probabilité de tomber dans le domaine D est le volume du cylindre C limité en haut par la surface de répartition et dont la base est le domaine D (fig. 8.3.5).

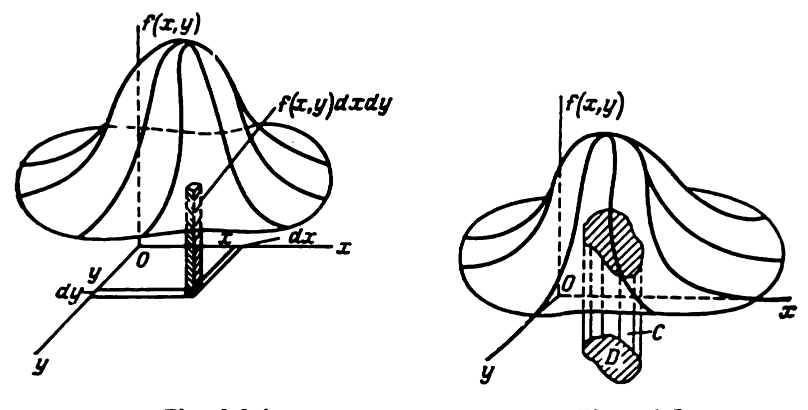


Fig. 8.3.4

Fig. 8.3.5

De la formule générale (8.3.3) on peut déduire celle concernant le rectangle R limité par les abscisses α et β et les ordonnées γ et δ (fig. 8.2.5):

$$P((X, Y) \subset R) = \int_{\alpha}^{\beta} \int_{\gamma}^{\delta} f(x, y)\, dx\, dy. \qquad (8.3.4)$$

Nous allons utiliser la formule (8.3.4) pour exprimer la fonction de répartition $F(x, y)$ d'un système en fonction de la densité de probabilité $f(x, y)$. La fonction de répartition $F(x, y)$ est la probabilité de tomber dans un quadrant infini; on peut se représenter ce dernier comme le rectangle limité par les abscisses $-\infty$ et x et les ordonnées $-\infty$ et y. En vertu de la formule (8.3.4) on a:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\, dx\, dy. \qquad (8.3.5)$$

Il est facile de voir que la densité de probabilité est douée des propriétés suivantes:

1. La densité de probabilité d'un système est une fonction non négative:

$$f(x, y) \geqslant 0.$$

Ceci est évident car la densité de probabilité est la limite du rapport de deux grandeurs non négatives: de la probabilité de tomber dans un rectangle et de l'aire de ce rectangle; par conséquent elle ne peut être négative.

2. L'intégrale double dans des limites infinies de la densité de probabilité du système est égale à l'unité:

$$\int_{-\infty}^{\infty}\int f(x, y)\,dx\,dy = 1. \qquad (8.3.6)$$

En effet l'intégrale (8.3.6) n'est rien d'autre que la probabilité de tomber dans le plan xOy entier, c'est-à-dire la probabilité d'un événement certain.

Du point de vue géométrique ceci signifie que le volume total d'un corps limité par la surface de répartition et le plan xOy est égal à l'unité.

Exemple 1. Soit un système de deux variables aléatoires (X, Y) suivant une loi de répartition de densité de probabilité

$$f(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}.$$

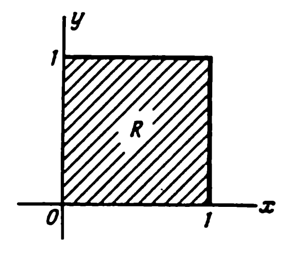


Fig. 8.3.6

Trouver la fonction de répartition $F(x, y)$. Trouver la probabilité pour le point aléatoire (X, Y) de se trouver dans le rectangle R (fig. 8.3.6).

Solution. La fonction de répartition $F(x, y)$ est donnée par la formule (8.3.5):

$$F(x, y) = \frac{1}{\pi^2}\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{dx\,dy}{(1+x^2)(1+y^2)} =$$

$$= \left(\frac{1}{\pi}\operatorname{arc\,tg} x + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arc\,tg} y + \frac{1}{2}\right).$$

La probabilité de tomber dans le rectangle R est donnée par la formule (8.3.4):

$$P((X, Y) \subset R) = \frac{1}{\pi^2}\int_0^1\int_0^1\frac{dx\,dy}{(1+x^2)(1+y^2)} =$$

$$= \frac{1}{\pi^2}\int_0^1\frac{dx}{1+x^2}\int_0^1\frac{1}{1+y^2} = \frac{1}{16}.$$

Exemple 2. La surface de répartition d'un système (X, Y) est un cône circulaire droit, ayant pour base un cercle de rayon R de centre à l'origine des

coordonnées. Trouver l'expression de la densité de probabilité. Trouver la probabilité pour que le point aléatoire (x, y) se trouve à l'intérieur d'un cercle C de rayon a, $a < R$, (fig. 8.3.7).

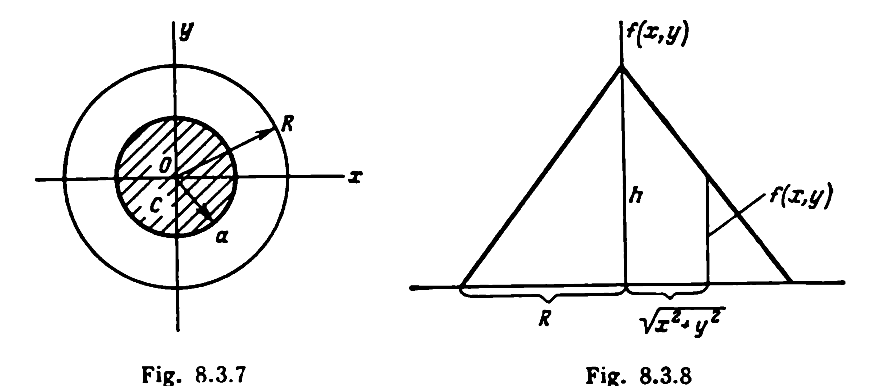


Fig. 8.3.7 Fig. 8.3.8

S o l u t i o n. Dans les notations de la figure 8.3.8 l'expression de la densité de probabilité à l'intérieur du cercle C s'écrit:

$$f(x, y) = \frac{h}{R}\left(R - \sqrt{x^2 + y^2}\right).$$

où h est la hauteur du cône. On choisit h de telle sorte que le volume du cône soit égal à l'unité, soit $\frac{1}{3}\pi R^2 h = 1$, d'où

$$h = \frac{3}{\pi R^2}$$

et

$$f(x, y) = \frac{3}{\pi R^3}\left(R - \sqrt{x^2 + y^2}\right).$$

La probabilité de tomber à l'intérieur du cercle C est donnée par la formule (8.3.4):

$$P((X, Y) \subset C) = \iint_{(C)} f(x, y)\, dx\, dy. \qquad (8.3.7)$$

Pour calculer l'intégrale (8.3.7) il est commode de passer aux coordonnées polaires r, φ:

$$P((X, Y) \subset C) = \int_0^a \int_0^{2\pi} \frac{3}{\pi R^3}(R - r)\, r\, dr\, d\varphi =$$

$$= \frac{6}{R^3}\int_0^a (R - r)\, r\, dr = 3\left(\frac{a}{R}\right)^2 - 2\left(\frac{a}{R}\right)^3.$$

8.4. Lois de répartition des variables d'un système. Lois de répartition conditionnelles

Connaissant la loi de répartition d'un système de deux variables aléatoires on peut toujours trouver celle des composantes. Dans le paragraphe 8.2 nous avons exprimé les fonctions de répartition des variables d'un système à l'aide de la fonction de répartition du système, plus exactement nous avons montré que:

$$F_1(x) = F(x, \infty); \quad F_2(y) = F(\infty, y). \tag{8.4.1}$$

Nous allons maintenant exprimer la densité de probabilité de chacune des variables d'un système en fonction de la densité de probabilité du système. En utilisant la formule (8.3.5) donnant la fonction de répartition en fonction de la densité de probabilité on a:

$$F_1(x) = F(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy,$$

d'où en dérivant par rapport à x on obtient l'expression de la densité de probabilité de X:

$$f_1(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x, y)\, dy. \tag{8.4.2}$$

D'une manière analogue on a:

$$f_2(y) = F_2'(y) = \int_{-\infty}^{\infty} f(x, y)\, dx. \tag{8.4.3}$$

Ainsi, pour obtenir la densité de probabilité de l'une des variables du système, il y a lieu d'intégrer la densité de probabilité du système dans des limites infinies sur l'argument correspondant à l'autre variable.

Les formules (8.4.1), (8.4.2) et (8.4.3) permettent, connaissant la loi de répartition du système (sa fonction de répartition ou sa densité de probabilité) de trouver la loi de répartition des variables faisant partie du système. On peut alors tout naturellement se poser le problème inverse qui consiste, connaissant les lois de répartition des variables d'un système, à trouver la loi de répartition du système. Il se trouve que dans le cas général c'est impossible: la connaissance des lois de répartition des variables du système ne suffit pas toujours à trouver la loi de répartition du système. Pour caractériser complètement un système il faut encore connaître la relation existant entre les variables du système. Cette relation peut être exprimée à l'aide des *lois de répartition conditionnelles*.

On appelle loi de répartition conditionnelle de la variable X du système (X, Y) la loi de répartition calculée sous la condition que l'autre variable Y a pris une certaine valeur y.

On peut donner une loi de répartition conditionnelle tant par la fonction de répartition que par la densité. On désigne par $F(x \mid y)$ la fonction de répartition conditionnelle, et par $f(x \mid y)$ la densité de probabilité conditionnelle. Vu le grand intérêt pratique que présentent les systèmes de variables continues, nous nous limitons dans ce cours à l'étude des lois conditionnelles données pour des densités de probabilité. Pour préciser la notion de loi de répartition conditionnelle nous allons étudier un exemple. Un système est formé de variables aléatoires L et Q qui sont la grandeur et le poids d'une race d'animaux. Supposons que nous nous intéressions à la grandeur L indépendamment du poids; c'est une variable aléatoire répartie suivant une certaine loi de densité de probabilité $f_1(l)$. Comme toujours, on peut étudier cette loi de probabilité en envisageant tous les animaux sans exception, enregistrant la grandeur de chacun d'eux et en traitant les données obtenues tout comme dans le cas d'une seule variable aléatoire. La loi $f_1(l)$ est la loi de répartition inconditionnelle de la grandeur des animaux. Cependant on peut également s'intéresser à la loi de répartition de grandeur des animaux d'un certain poids, par exemple de 2 kg. Pour trouver cette loi nous allons mesurer la grandeur non pas de tous les animaux, mais seulement de ceux dont le poids est voisin de 2 kg; nous obtiendrons alors la loi de répartition conditionnelle de la grandeur des animaux $f(l \mid q)$ pour $q = 2$. Cette loi conditionnelle diffère en général de la loi inconditionnelle $f_1(l)$; il est évident que les animaux plus lourds seront en moyenne plus grands, par conséquent, la loi de répartition conditionnelle de grandeur d'un animal $f(l \mid q)$ dépend de la valeur donnée de q.

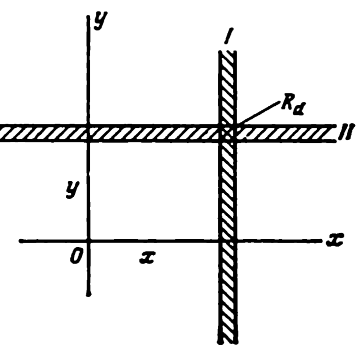


Fig. 8.4.1

Connaissant la loi de répartition de l'une des variables du système et la loi de répartition conditionnelle de la seconde on peut déterminer la loi de répartition du système. Nous allons trouver la formule exprimant cette relation pour les variables aléatoires continues. A cet effet nous utiliserons la notion d'élément de probabilité. Soit le rectangle élémentaire R_d de côtés dx et dy adjacent au point (x, y) (fig. 8.4.1). La probabilité de tomber dans ce rectangle, donnée par l'élément de probabilité $f(x, y)\,dx\,dy$, est égale à la probabilité pour le point aléatoire (X, Y) de se trouver simultanément dans la bande élémentaire I de base dx et dans la bande II de base dy:

$$f(x, y)\,dx\,dy = P((X, Y) \subset R_d) =$$
$$= P((x < X < x + dx)(y < Y < y + dy)).$$

La probabilité du produit de ces deux événements est en vertu du théorème de multiplication des probabilités égale à la probabilité de tomber dans la bande élémentaire I, multipliée par la probabilité conditionnelle de tomber dans la bande élémentaire II, cette dernière probabilité calculée sous la condition que le premier événement ait été réalisé. A la limite cette condition équivaut à la condition $X = x$; par conséquent:

$$f(x, y)\,dx\,dy = f_1(x)\,dx\,f(y \mid x)\,dy,$$

d'où:

$$f(x, y) = f_1(x)\,f(y \mid x), \tag{8.4.4}$$

c'est-à-dire que la densité de probabilité d'un système de deux variables est égale à la densité de probabilité de l'une d'elles multipliée par la densité de probabilité conditionnelle de l'autre, calculée sous la condition que la première ait une valeur donnée.

La formule (8.4.4) est souvent appelée *théorème de multiplication des lois de répartition*. Ce théorème est équivalent à celui de multiplication des probabilités pour les événements.

Il est évident que la formule (8.4.4) peut prendre une autre forme si ce n'est pas la variable X mais Y qui est donnée:

$$f(x, y) = f_2(y)\,f(x \mid y). \tag{8.4.5}$$

A partir des équations (8.4.4) et (8.4.5) on trouve les expressions des probabilités conditionnelles $f(y \mid x)$ et $f(x \mid y)$ en fonction des lois inconditionnelles:

$$\left.\begin{aligned} f(y \mid x) &= \frac{f(x, y)}{f_1(x)}, \\ f(x \mid y) &= \frac{f(x, y)}{f_2(y)}, \end{aligned}\right\} \tag{8.4.6}$$

ou, compte tenu des formules (8.4.2) et (8.4.3),

$$\left.\begin{aligned} f(y \mid x) &= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\,dy}; \\ f(x \mid y) &= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\,dx}. \end{aligned}\right\} \tag{8.4.7}$$

8.5. Variables aléatoires liées et indépendantes

Lors de l'étude des systèmes de variables aléatoires il y a toujours lieu de tenir compte du degré et du caractère de leur dépendance mutuelle qui peut être plus ou moins prononcée, plus ou moins importante. Dans certains cas les variables aléatoires peuvent être si étroitement

liées que connaissant la valeur de l'une d'elles on peut déterminer exactement la valeur de l'autre. Dans l'autre cas limite les variables aléatoires sont si faiblement liées que l'on peut pratiquement les considérer comme indépendantes.

La notion de variable aléatoire indépendante est l'une des notions fondamentales de la théorie des probabilités.

La variable aléatoire Y est dite indépendante de la variable aléatoire X si la loi de répartition de Y ne dépend pas de la valeur prise par X.

Dans le cas continu, la condition d'indépendance des variables aléatoires X et Y s'écrit:

$$f(y \mid x) = f_2(y)$$

quel que soit y.

Au contraire, si Y dépend de X on a:

$$f(y \mid x) \neq f_2(y).$$

Nous allons montrer que la *dépendance ou l'indépendance des variables aléatoires est une propriété mutuelle*, c'est-à-dire si Y ne dépend pas de X, X ne dépend pas non plus de Y.

En effet, supposons que Y ne dépende pas de X:

$$f(y \mid x) = f_2(y). \qquad (8.5.1)$$

En vertu des formules (8.4.4) et (8.4.5) on a alors:

$$f_1(x) f(y \mid x) = f_2(y) f(x \mid y),$$

d'où, compte tenu de (8.5.1),

$$f(x \mid y) = f_1(x),$$

ce qu'il fallait démontrer.

Comme les propriétés de dépendance et d'indépendance des variables aléatoires sont mutuelles, on peut donner une nouvelle définition des variables aléatoires indépendantes.

Les variables aléatoires X et Y sont dites indépendantes si la loi de répartition de chacune d'elles ne dépend pas des valeurs prises par l'autre. Dans le cas contraire les variables aléatoires X et Y sont dites liées.

Pour les variables aléatoires continues le théorème de multiplication des lois de répartition s'écrit:

$$f(x, y) = f_1(x) f_2(y), \qquad (8.5.2)$$

c'est-à-dire que *la densité de probabilité d'un système de variables aléatoires indépendantes est égale au produit des densités des variables faisant partie du système.*

La condition (8.5.2) peut être considérée comme la condition nécessaire et suffisante d'indépendance des variables aléatoires.

Souvent la forme même de la fonction $f(x, y)$ permet de voir que les variables X, Y sont indépendantes: si la densité de probabilité $f(x, y)$ peut s'écrire comme le produit de deux fonctions dont l'une dépend seulement de x et l'autre seulement de y, les variables aléatoires sont indépendantes.

E x e m p l e. La densité de probabilité du système (X, Y) est de la forme:

$$f(x, y)=\frac{1}{\pi^2(x^2+y^2+x^2y^2+1)}.$$

Déterminer si les variables aléatoires X et Y sont liées ou non.

S o l u t i o n. En décomposant le dénominateur en produit de facteurs on a:

$$f(x, y)=\frac{1}{\pi(1+x^2)}\frac{1}{\pi(1+y^2)}.$$

Le fait que la fonction $f(x, y)$ se décompose en un produit de deux fonctions dont l'une dépend seulement de x et l'autre seulement de y permet de conclure que les variables X et Y doivent être indépendantes. En effet, en vertu des formules (8.4.2) et (8.4.3) on a:

$$f_1(x)=\frac{1}{\pi(1+x^2)}\int_{-\infty}^{\infty}\frac{dy}{\pi(1+y^2)}=\frac{1}{\pi(1+x^2)};$$

et d'une manière analogue:

$$f_2(y)=\frac{1}{\pi(1+y^2)},$$

d'où

$$f(x, y) = f_1(x) f_2(y),$$

ce qui signifie que X et Y sont indépendantes.

Le critère mentionné pour juger si des variables aléatoires sont liées ou non n'est applicable que si la loi de répartition du système est connue. Le plus souvent dans la pratique on ne connaît pas la loi du système (X, Y), mais seulement les lois des composantes, et l'on a toutes les raisons de penser qu'elles sont indépendantes. On peut alors écrire la densité de probabilité du système comme le produit des densités des composantes.

Nous allons nous arrêter un peu plus en détail sur les notions fort importantes de « dépendance » ou d'« indépendance» des variables aléatoires.

La notion de « dépendance » de variables aléatoires utilisée en théorie des probabilités diffère quelque peu de la notion usuelle de « dépendance » des grandeurs employée en mathématiques. En effet, on entend habituellement par dépendance des grandeurs un seul type de dépendance, à savoir la dépendance totale, dite *fonctionnelle*. Deux grandeurs X et Y sont fonctionnellement dépendantes si connaissant la valeur de l'une d'elles on peut trouver exactement la valeur de l'autre.

En théorie des probabilités cette notion est comprise dans un sens plus général, on parle d'une dépendance *probabiliste* ou *stochastique*. Si deux variables Y et X sont liées par une relation probabiliste, connaissant la valeur de X on ne peut pas déterminer la valeur exacte de Y, mais seulement donner sa loi de répartition qui dépend de la valeur de X.

La dépendance probabiliste peut être plus ou moins forte; plus cette dépendance est forte, plus elle se rapproche d'une relation fonctionnelle. Ainsi, une relation fonctionnelle peut être considérée comme le cas limite de la dépendance probabiliste forte. L'autre cas limite est l'indépendance totale des variables aléatoires. Toutes les variétés de la dépendance probabiliste, de la plus forte à la plus faible, se trouvent entre ces deux cas limites. Les grandeurs physiques entre lesquelles nous voyons une relation fonctionnelle sont en réalité liées par une relation probabiliste bien étroite, c'est-à-dire que pour une valeur donnée de l'une de ces grandeurs l'autre varie dans des limites si étroites qu'on peut pratiquement la considérer comme une grandeur certaine. D'autre part, les grandeurs que nous considérons dans la pratique indépendantes, sont en réalité mutuellement liées, mais cette liaison est si faible qu'on peut la négliger.

Dans la pratique on rencontre fréquemment des relations probabilistes entre des variables aléatoires. Dire que les variables aléatoires X et Y sont liées par une relation probabiliste ne signifie pas que les variations de X entraînent des variations déterminées de Y; ceci signifie seulement que lorsque X varie, Y tend également à varier (par exemple à croître ou décroître lorsque X croît). Cette tendance ne se manifeste qu'en moyenne, grossièrement, et peut ne pas toujours se réaliser.

Considérons par exemple deux variables aléatoires X et Y qui sont la taille d'un homme pris au hasard et son poids. Il est évident qu'entre les variables X et Y il y a une relation probabiliste, car dans le cas général les hommes plus grands sont plus lourds. On peut même trouver une formule empirique remplaçant cette relation probabiliste par une relation fonctionnelle. Citons en guise d'exemple la formule bien connue donnant la relation approchée entre la taille et le poids:

$$Y\ (\mathrm{kg}) = X\ (\mathrm{cm}) - 100.$$

Il est évident que les formules de ce genre ne sont pas exactes et n'expriment une relation qu'en moyenne, une tendance générale avec des écarts possibles dans chaque cas particulier. Considérons maintenant les deux variables aléatoires suivantes: la taille X d'un homme pris au hasard et son âge Z. Il est clair que pour des adultes les variables X et Y sont pratiquement indépendantes; au contraire pour un enfant elles sont liées.

Nous allons encore donner quelques exemples de variables aléatoires plus ou moins liées entre elles.

1. On prend au hasard une pierre dans un cailloutis. Le poids Q d'un caillou et sa dimension maximale L sont des variables aléatoires, liées de toute évidence par une relation probabiliste forte.

2. Des fusées sont lancées dans une région donnée de l'océan. ΔX est l'erreur longitudinale du point d'impact et ΔV l'erreur de vitesse de la fusée à la fin du vol. Les variables ΔX et ΔV sont de toute évidence liées, car l'erreur ΔV est l'une des causes principales de l'erreur longitudinale ΔX.

3. Un avion en vol détermine sa hauteur au-dessus de la surface de la terre à l'aide d'un appareil barométrique. Considérons les deux variables aléatoires suivantes: ΔH, l'erreur de mesure de l'altitude et G, le poids du combustible restant dans le réservoir au moment de la mesure. Les variables ΔH et G sont pratiquement indépendantes.

Dans le paragraphe suivant nous allons étudier certaines caractéristiques numériques des systèmes des variables aléatoires permettant d'estimer le degré de dépendance des variables.

8.6. Caractéristiques numériques des systèmes de deux variables aléatoires. Covariance. Coefficient de corrélation

Dans le chapitre 5 nous avons étudié les caractéristiques numériques d'une variable aléatoire X, à savoir les moments initiaux et centrés de différents ordres. Deux de ces caractéristiques jouent un rôle particulièrement important, ce sont l'espérance mathématique m_x et la variance D_x.

On peut introduire des caractéristiques numériques semblables pour un système de deux variables aléatoires.

On appelle *moment initial* d'ordre k, s du système (X, Y) l'espérance mathématique du produit de X^k par Y^s:

$$\alpha_{k,s} = M[X^k Y^s]. \tag{8.6.1}$$

On appelle *moment centré* d'ordre k, s du système (X, Y) l'espérance mathématique du produit des puissances k et s des moments centrés correspondants, soit:

$$\mu_{k,s} = M[\mathring{X}^k \mathring{Y}^s], \tag{8.6.2}$$

où $\mathring{X} = X - m_x$, $\mathring{Y} = Y - m_y$.

Ecrivons les formules permetttant de calculer directement les moments. Pour les variables aléatoires discrètes on a:

$$\alpha_{k,s} = \sum_i \sum_j x_i^k y_j^s p_{ij}, \tag{8.6.3}$$

$$\mu_{k,s} = \sum_i \sum_j (x_i - m_x)^k (y_j - m_y)^s p_{ij}, \tag{8.6.4}$$

où $p_{ij} = P((X = x_i)(Y = y_j))$ est la probabilité pour que le système (X, Y) prenne la valeur (x_i, y_j), la sommation s'étendant à toutes les valeurs possibles des variables X, Y.

Pour les variables aléatoires continues on a:

$$\alpha_{k,s} = \iint_{-\infty}^{\infty} x^k y^s f(x, y)\, dx\, dy, \tag{8.6.5}$$

$$\mu_{k,s} = \iint_{-\infty}^{\infty} (x - m_x)^k (y - m_y)^s f(x, y)\, dx\, dy, \tag{8.6.6}$$

où $f(x, y)$ est la densité de probabilité du système.

En plus des nombres k et s caractérisant l'ordre des moments des variables isolément, on considère encore l'ordre total $k + s$, égal à la somme des puissances de X et de Y. Pour indiquer l'ordre du moment d'un système on donne son ordre total. Dans la pratique on utilise seulement les moments du premier et du second ordre.

Les moments initiaux d'ordre un sont les *espérances mathématiques* bien connues des variables X et Y du système, soit:

$$m_x = \alpha_{1,0} = M[X^1 Y^0] = M[X];$$
$$m_y = \alpha_{0,1} = M[X^0 Y^1] = M[Y].$$

L'ensemble des espérances mathématiques m_x, m_y est la *caractéristique de position* du système. Du point de vue géométrique ce sont les coordonnées du point moyen dans le plan autour duquel il y a dispersion du point (X, Y).

En plus des moments initiaux d'ordre un, dans les applications pratiques on utilise encore les *moments centrés deux* du système. Deux de ces moments sont les variances des variables X et Y que nous connaissons déjà, soit:

$$D_x = \mu_{2,0} = M[\overset{\circ}{X}{}^2 \overset{\circ}{Y}{}^0] = M[\overset{\circ}{X}{}^2] = D[X],$$
$$D_y = \mu_{0,2} = M[\overset{\circ}{X}{}^0 \overset{\circ}{Y}{}^2] = M[\overset{\circ}{Y}{}^2] = D[Y],$$

caractérisant la *dispersion* du point aléatoire dans la direction des axes Ox et Oy.

Le *moment central mixte deux* joue un rôle particulièrement important. C'est la quantité

$$\mu_{1,1} = M[\overset{\circ}{X}\overset{\circ}{Y}],$$

donc l'espérance mathématique du produit des variables centrées.

Vu son rôle important dans la théorie des systèmes de variables aléatoires, on lui donne une désignation particulière:

$$K_{xy} = M[\overset{\circ}{X}\overset{\circ}{Y}] = M[(X - m_x)(Y - m_y)], \tag{8.6.7}$$

et on l'appelle *covariance* des variables aléatoires X, Y.

Pour les variables aléatoires discrètes la covariance est donnée par la formule:

$$K_{xy}=\sum_i\sum_j (x_i-m_x)(y_j-m_y)\,p_{i,j}, \tag{8.6.8}$$

et pour les variables continues par la formule:

$$K_{xy}=\int\limits_{-\infty}^{\infty}\int (x-m_x)(y-m_y)\,f(x,\ y)\,dx\,dy. \tag{8.6.9}$$

Nous allons préciser le sens et le rôle de cette caractéristique.

La covariance est une caractéristique du système décrivant, en plus de la dispersion des variables X et Y, l e u r l i a i s o n. Pour nous en rendre compte nous allons montrer que la *covariance des variables aléatoires indépendantes est nulle*.

Nous démontrons cette proposition pour le cas continu *). Soient X, Y des variables continues, indépendantes, de densité de probabilité $f(x,\ y)$. Dans le § 8.5 nous avons démontré que pour des variables indépendantes on a:

$$f(x,\ y)=f_1(x)\,f_2(y), \tag{8.6.10}$$

où $f_1(x)$, $f_2(y)$ sont les densités de probabilité des variables X et Y respectivement.

En substituant (8.6.10) dans (8.6.9) on voit que l'intégrale (8.6.9) devient un produit de deux intégrales, en effet:

$$K_{xy}=\int\limits_{-\infty}^{\infty}(x-m_x)\,f_1(x)\,dx\int\limits_{-\infty}^{\infty}(y-m_y)\,f_2(y)\,dy.$$

L'intégrale

$$\int\limits_{-\infty}^{\infty}(x-m_x)\,f_1(x)\,dx$$

n'est rien d'autre que le premier moment centré de X et donc est égale à zéro; le second facteur est nul pour la même raison; par conséquent, pour des variables aléatoires indépendantes on a $K_{xy}=0$.

Ainsi, si la covariance de deux variables aléatoires est différente de zéro c'est un indice montrant qu'elles sont liées.

La formule (8.6.7) montre que la covariance caractérise non seulement la dépendance des variables aléatoires mais également leur dispersion. En effet si, par exemple, l'une des variables du système $(X,\ Y)$ s'écarte peu de son espérance mathématique (variable presque certaine), la covariance sera petite même si les variables X, Y sont étroitement liées entre elles. Alors pour caractériser seulement la liai-

*) Pour les variables discrètes la démonstration est analogue.

son des variables X, Y on passe du moment K_{xy} à la grandeur sans dimension

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}, \tag{8.6.11}$$

où σ_x, σ_y sont les écarts quadratiques moyens des variables X, Y. Cette caractéristique s'appelle *coefficient de corrélation* des variables X et Y. Il est évident que le coefficient de corrélation s'annule avec la covariance; par conséquent le *coefficient de corrélation des variables aléatoires indépendantes est égal à zéro.*

On appelle *non corrélées* les variables aléatoires dont la covariance (donc également le coefficient de corrélation) est nulle.

Nous allons voir maintenant si les notions de « non-corrélation » et d'« indépendance » sont équivalentes. Nous avons démontré ci-dessus que deux variables aléatoires indépendantes sont toujours non corrélées. Il faut voir maintenant si l'assertion inverse selon laquelle deux variables non corrélées sont indépendantes est vraie. Il se trouve que non. On peut trouver des exemples de variables aléatoires non corrélées mais liées. La nullité du coefficient de corrélation des variables aléatoires est une condition nécessaire mais non suffisante de leur indépendance. L'indépendance des variables aléatoires entraîne leur non-corrélation, l'inverse n'étant pas vrai, la non-corrélation n'entraînant pas l'indépendance. La condition d'indépendance des variables aléatoires est une condition plus rigoureuse que celle de non-corrélation.

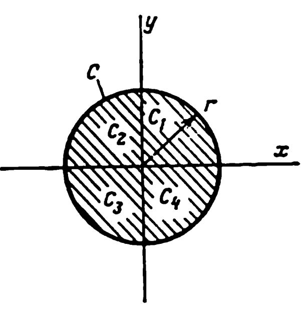


Fig. 8.6.1

Nous allons voir ce qui vient d'être dit sur un exemple. Soit un système de variables aléatoires (X, Y) uniformément réparties à l'intérieur d'un cercle C de rayon r et de centre à l'origine des coordonnées (fig. 8.6.1).

La densité de probabilité du système (X, Y) s'écrit:

$$f(x, y) = \begin{cases} c \text{ pour } x^2 + y^2 < r^2, \\ 0 \text{ pour } x^2 + y^2 > r^2. \end{cases}$$

La condition $\int\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int\int_{(C)} c\, dx\, dy = 1$ donne $c = 1/\pi r^2$.

Il est facile de voir que dans cet exemple les variables sont dépendantes. En effet, il est immédiat que si par exemple $X = 0$, Y peut, avec une même probabilité, prendre toutes les valeurs comprises entre $-r$ et $+r$; mais si $X = r$, Y ne peut prendre qu'une seule valeur égale à zéro; dans le cas général la gamme des valeurs possibles de Y dépend de la valeur de X.

Pour voir si ces variables sont corrélées, calculons leur covariance. En vertu de la symétrie des conditions du problème, $m_x = m_y = 0$ et on a:

$$K_{xy} = \iint\limits_{(C)} xy\,f(x, y)\,dx\,dy = \frac{1}{\pi r^2} \iint\limits_{(C)} xy\,dx\,dy. \qquad (8.6.12)$$

Pour calculer cette intégrale on divise le domaine d'intégration (cercle C) en quatre secteurs C_1, C_2, C_3, C_4 correspondant aux quatre quadrants. Dans les secteurs C_1 et C_3 la fonction sous l'intégrale est positive, dans les secteurs C_2 et C_4 elle est négative; de plus, les intégrales dans tous ces secteurs sont égales en valeur absolue; par conséquent l'intégrale (8.6.12) est nulle et les variables X, Y ne sont pas corrélées.

On voit ainsi que des variables aléatoires non corrélées ne sont pas forcément indépendantes.

Le coefficient de corrélation caractérise non pas une relation quelconque mais seulement une relation dite *linéaire*. On dit que deux variables aléatoires sont liées par une relation probabiliste linéaire si lorsque l'une d'elles croît, l'autre accuse une tendance à croître (à décroître) suivant une loi linéaire. Cette tendance linéaire peut être plus ou moins prononcée et s'approcher plus ou moins d'une relation fonctionnelle, c'est-à-dire d'une relation linéaire au sens strict. Le coefficient de corrélation indique à quel point la relation linéaire entre deux variables aléatoires est intime. Si les variables aléatoires X et Y sont liées par une relation fonctionnelle exacte

$$Y = aX + b,$$

on a $r_{xy} = \pm 1$, le signe «+» ou «—» correspond au signe du coefficient a. Dans le cas général, lorsque les variables X et Y sont liées par une relation probabiliste arbitraire, le coefficient de corrélation peut se trouver dans les limites:

$$-1 < r_{xy} < 1\ ^{*}).$$

Quand $r_{xy} > 0$, la corrélation est dite *positive*, dans le cas $r_{xy} < 0$ elle est dite *négative*. Une corrélation positive signifie que lorsque l'une des variables croît, l'autre a également tendance à croître en moyenne; une corrélation négative signifie que lorsque l'une des variables croît, l'autre tend en moyenne à décroître.

Dans l'exemple envisagé d'un système de deux variables aléatoires (X, Y) uniformément réparties à l'intérieur d'un cercle, bien qu'entre X et Y il existe une relation, celle-ci n'est pas linéaire; lorsque X croît, seule la gamme des variations de Y change, sa valeur *moyenne* restant la même, donc X et Y sont non corrélées.

---

*) Nous allons démontrer cette assertion plus loin, dans le § 10.3, après l'introduction de certains théorèmes de la théorie des probabilités permettant de simplifier la démonstration.

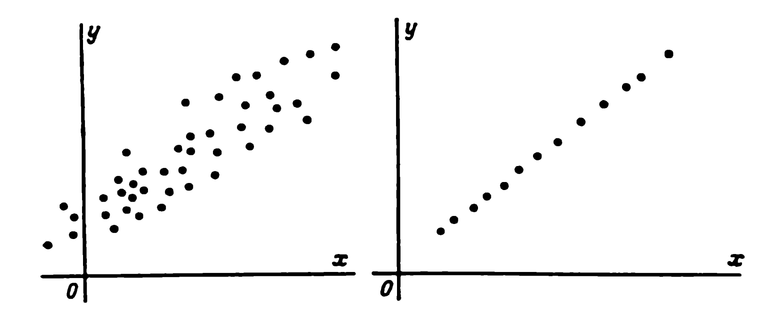


Fig. 8.6.2

Fig. 8.6.3

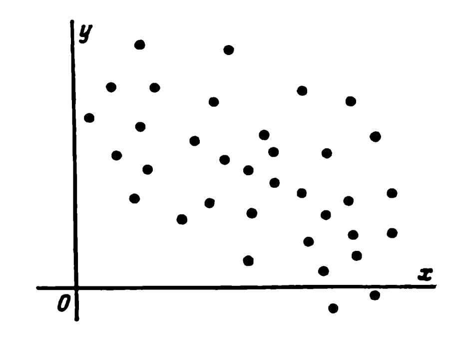


Fig. 8.6.4

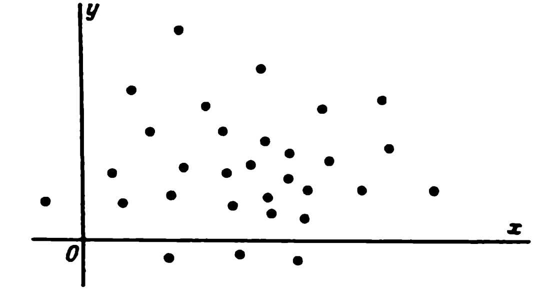


Fig. 8.6.5

Nous allons donner quelques exemples de variables aléatoires à corrélation positive et négative.

1. Le poids et la taille d'un homme ont une corrélation positive.

2. Le temps de réglage d'un appareil avant sa mise en service et le temps de fonctionnement sans défaillance ont une corrélation positive (si, bien entendu, les pertes de temps sont rationnelles). Inversement, entre le temps de préparation et le nombre de défauts décelés lors du fonctionnement de l'appareil il y a une corrélation négative.

3. Dans un tir par rafales, les coordonnées des points d'impact des différentes balles sont liées par une corrélation positive (car il y a des erreurs de visée communes à tous les coups).

4. Deux coups sont tirés sur une cible l'un après l'autre ; le point d'impact du premier coup est enregistré et une correction de visée est introduite proportionnelle à l'erreur du premier coup et de signe inverse. Les coordonnées des points d'impact du premier et du second coup seront dans ce cas liées par une corrélation négative.

Si nous avons à notre disposition les résultats d'une série d'expériences effectuées sur un système de deux variables aléatoires (X, Y), nous pouvons juger grossièrement de l'existence d'une corrélation plus ou moins importante entre ces variables d'après le graphique où sont portés les points correspondant à tous les couples observés de leurs valeurs. Par exemple, si les points se disposent comme indiqué sur la figure 8.6.2 il y a une corrélation positive nette entre ces variables. Cette corrélation positive est encore plus nette sur la figure 8.6.3, où elle est voisine d'une relation fonctionnelle. Sur la figure 8.6.4 on peut voir un cas d'une corrélation négative assez faible. Enfin sur la figure 8.6.5 on a représenté un cas de variables aléatoires non corrélées. Dans les applications pratiques, il est toujours bon de porter les couples observés sur un graphique pour se faire une idée qualitative du type de corrélation.

Dans le chapitre 14 nous étudierons les méthodes de détermination des caractéristiques des systèmes de variables aléatoires à partir des données expérimentales.

8.7. Système d'un nombre arbitraire de variables aléatoires

Dans la pratique on rencontre souvent des systèmes composés de plus de deux variables aléatoires. On peut interpréter ces systèmes comme des points aléatoires ou des vecteurs aléatoires dans l'espace de dimension respective.

Citons d'abord quelques exemples.

1. Le point d'explosion d'un projectile fusant dans l'espace est caractérisé par trois coordonnées cartésiennes (X, Y, Z) ou par trois coordonnées sphériques (R, Φ, Θ).

2. L'ensemble de n mesures successives d'une grandeur variable X est un système de n variables aléatoires $(X_1, X_2, \ldots, X_n)$.
3. On tire une rafale de n coups. L'ensemble des coordonnées des n points d'impact dans le plan forme un système de $2n$ variables aléatoires (abscisses et ordonnées des points d'impact):

$$(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n).$$

4. La vitesse initiale d'un éclat est un vecteur aléatoire caractérisé par trois grandeurs aléatoires: la valeur de la vitesse V_0 et les deux angles Φ et Θ déterminant la direction de vol de l'éclat dans un système de coordonnées sphériques.

Un système d'un nombre arbitraire de variables aléatoires est entièrement caractérisé par la loi de répartition du système pouvant être donnée par la fonction de répartition ou la densité de probabilité.

On appelle fonction de répartition d'un système de n variables aléatoires $(X_1, X_2, \ldots, X_n)$ la probabilité pour que soient vérifiées simultanément n inégalités du type $X_i < x_i$:

$$F(x_1, x_2, \ldots, x_n) = P((X_1 < x_1)(X_2 < x_2) \ldots (X_n < x_n)). \tag{8.7.1}$$

On appelle densité de probabilité d'un système de n variables aléatoires continues la dérivée n-ième partielle mixte de la fonction $F(x_1, x_2, \ldots, x_n)$ prise une fois par rapport à chaque argument:

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}. \tag{8.7.2}$$

Connaissant la loi de répartition du système, on peut trouver celle des composantes. La fonction de répartition de chacune des variables du système s'obtient en posant tous les autres arguments dans la fonction de répartition du système égaux à ∞:

$$F_1(x_1) = F(x_1, \infty, \ldots, \infty). \tag{8.7.3}$$

Si $(X_1, X_2, \ldots, X_k)$ est un sous-système du système $(X_1, X_2, \ldots, X_n)$, la fonction de répartition de ce sous-système est donnée par la formule:

$$F_{1,2,\ldots k}(x_1, x_2, \ldots, x_k) = F(x_1, x_2, \ldots, x_k, \infty, \ldots, \infty). \tag{8.7.4}$$

Pour obtenir la densité de probabilité de chacune des variables du système il y a lieu d'intégrer la densité de probabilité du système dans des limites infinies par rapport à tous les autres arguments:

$$f_1(x_1) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2 \ldots dx_n. \tag{8.7.5}$$

La densité de probabilité du sous-système $(X_1, X_2, \ldots, X_k)$ issu du système $(X_1, X_2, \ldots, X_n)$ est égale à

$$f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_{k+1} \ldots dx_n. \qquad (8.7.6)$$

On appelle loi de répartition conditionnelle du sous-système $(X_1, X_2, \ldots, X_k)$ sa loi de répartition calculée sous la condition que les variables $X_{k+1}, \ldots, X_n$ restantes aient pris les valeurs $x_{k+1}, \ldots, x_n$.

La densité de probabilité conditionnelle est donnée par la formule:

$$f(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{f_{k+1,\ldots,n}(x_{k+1}, \ldots, x_n)}. \qquad (8.7.7)$$

Les variables aléatoires $(X_1, X_2, \ldots, X_n)$ sont dites *indépendantes* si la loi de répartition de tout sous-système appartenant au système $(X_1, X_2, \ldots, X_n)$ ne dépend pas des valeurs prises par les autres variables du système.

La densité de probabilité d'un système de variables aléatoires indépendantes est égale au produit des densités de toutes les variables du système:

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n). \qquad (8.7.8)$$

La probabilité pour le point aléatoire $(X_1, X_2, \ldots, X_n)$ de se trouver dans un domaine D de dimension n s'exprime par l'intégrale à n dimensions:

$$P((X_1, X_2, \ldots, X_n) \subset D) = \int \ldots \int_{(D)} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n. \qquad (8.7.9)$$

La formule (8.7.9) est la formule de base pour calculer les probabilités des événements ne se ramenant pas au « schéma du tirage exhaustif ». En effet, si l'événement A nous intéressant ne se ramène pas au cas du « tirage exhaustif », sa probabilité ne peut pas être calculée directement. Si de plus il n'est pas possible de faire un nombre suffisant d'expériences semblables et de déterminer d'une manière approchée la probabilité de l'événement A d'après sa fréquence, on adopte le schéma de calcul suivant. On passe des événements aux variables aléatoires (le plus souvent continues) et l'on associe à l'événement A l'événement consistant en ce que le système de variables aléatoires $(X_1, X_2, \ldots, X_n)$ se trouve à l'intérieur d'un domaine D. Dans ce cas la probabilité de l'événement A peut être calculée par la formule (8.7.9).

E x e m p l e. Trois paramètres X, Y, Z d'un appareil sont soumis à des essais. Pour que l'appareil soit accepté il faut que les événements suivants soient simultanément réalisés:

1) $x_0 < X < x_1$;
2) $Y < X$;
3) $X > Z > Y$.

Les paramètres X, Y, Z sont trois variables aléatoires continues, à valeurs positives, de densité de probabilité $f(x, y, z)$. Trouver la probabilité pour que l'appareil soit accepté.

S o l u t i o n. En désignant par A l'événement consistant en ce que l'appareil soit accepté on trouve:

$$P(A)=\int_{x_0}^{x_1}\left\{\int_{0}^{x}\left\{\int_{x}^{y} f(x, y, z)\,dz\right\} dy\right\} dx.$$

8.8. Caractéristiques numériques d'un système de plusieurs variables aléatoires

La loi de répartition d'un système, donnée par la fonction de répartition ou la densité de probabilité, est une caractéristique complète du système. Cependant, il est rare que cette caractéristique complète puisse être utilisée. Parfois les données expérimentales trop pauvres ne permettent pas de trouver la loi de répartition du système. Dans d'autres cas l'emploi d'un appareil aussi compliqué que sont les lois de répartition ne se justifie pas par la précision assez médiocre exigée des résultats. Enfin, dans certains problèmes on connaît d'avance la forme de la loi de répartition (loi normale par exemple) et il y a lieu d'en trouver les paramètres.

Dans tous ces cas au lieu des lois de répartition on utilise une description approchée, incomplète, d'un système de variables aléatoires, ceci à l'aide d'un nombre restreint de caractéristiques numériques.

Enumérons ce minimum de caractéristiques permettant de caractériser un système de n variables aléatoires $X_1, X_2, \ldots, X_n$:

1) n espérances mathématiques

$$m_1, m_2, \ldots, m_n,$$

caractérisant les valeurs moyennes des variables;

2) n variances

$$D_1, D_2, \ldots, D_n,$$

caractérisant leur dispersion;

3) $n(n-1)$ covariances

$$K_{ij}=M[\overset{\circ}{X}_i\overset{\circ}{X}_j] \qquad (i \neq j),$$

où

$$\overset{\circ}{X}_i=X_i-m_i; \quad \overset{\circ}{X}_j=X_j-m_j,$$

caractérisant la corrélation de toutes les variables du système deux à deux.

Notons qu'en fait la variance de chacune des variables aléatoires X_i n'est rien d'autre qu'un cas particulier de la covariance, à savoir la covariance de la variable X_i et de la même variable X_i:

$$D_i = K_{ii} = M[\mathring{X}_i \mathring{X}_i] = M[\mathring{X}_i^2].$$

Il est commode de disposer toutes les covariances et les variances dans un tableau rectangulaire (appelé *matrice*)

$$\begin{Vmatrix} K_{11} & K_{12} & \ldots & K_{1n} \\ K_{21} & K_{22} & \ldots & K_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ K_{n1} & K_{n2} & \ldots & K_{nn} \end{Vmatrix}.$$

Ce tableau est appelé *matrice de corrélation* du système de variables aléatoires $(X_1, X_2, \ldots, X_n)$.

Il est évident que tous les éléments de la matrice de corrélation ne sont pas différents. Par définition de la covariance on a $K_{ij} = K_{ji}$, c'est-à-dire que les *éléments de la matrice de corrélation symétriques par rapport à la diagonale principale sont égaux entre eux*. C'est pour cela que souvent on n'écrit pas toute la matrice de corrélation mais seulement sa moitié, y compris la diagonale principale:

$$\begin{Vmatrix} K_{11} & K_{12} & \ldots & K_{1n} \\ & K_{22} & \ldots & K_{2n} \\ & \cdot & \cdot & \cdot \\ & & & K_{nn} \end{Vmatrix}.$$

Pour simplifier l'écriture on désigne souvent la matrice de corrélation d'éléments K_{ij} par le symbole $\| K_{ij} \|$.

Dans la diagonale principale de la matrice on trouve les variances des variables aléatoires $X_1, X_2, \ldots, X_n$.

Dans le cas des variables non corrélées, tous les éléments de la matrice de corrélation sauf ceux de diagonale sont nuls:

$$\| K_{ij} \| = \begin{Vmatrix} D_1 & 0 & 0 & \ldots & 0 \\ & D_2 & 0 & \ldots & 0 \\ & & D_3 & \ldots & 0 \\ & & & \cdot \cdot \cdot & \cdot \\ & & & & D_n \end{Vmatrix}.$$

Une telle matrice est dite *diagonale*.

Pour mieux faire apparaître la corrélation des variables aléatoires, indépendamment de leur dispersion, au lien de la matrice de

corrélation $\| K_{ij} \|$ on utilise souvent la *matrice de corrélation normée* $\| r_{ij} \|$ dont les éléments sont non pas les covariances mais les coefficients de corrélation :

$$r_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j},$$

où

$$\sigma_i = \sqrt{D_i}, \quad \sigma_j = \sqrt{D_j}.$$

Il est évident que tous les éléments diagonaux de cette matrice sont égaux à l'unité. La matrice de corrélation normée est de la forme :

$$\| r_{ij} \| = \left\| \begin{matrix} 1 & r_{12} & r_{13} & \ldots & r_{1n} \\ & 1 & r_{23} & \ldots & r_{2n} \\ & & 1 & \ldots & r_{3n} \\ & & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ & & & & 1 \end{matrix} \right\|$$

Introduisons la notion de systèmes non corrélés de variables aléatoires (ou de *vecteurs non corrélés*). Soient deux systèmes de variables aléatoires :

$$(X_1, X_2, \ldots, X_n); \quad (Y_1, Y_2, \ldots, Y_n)$$

ou deux vecteurs aléatoires dans l'espace à n dimensions : $\vec{X}$ de composantes $(X_1, X_2, \ldots, X_n)$ et $\vec{Y}$ de composantes $(Y_1, Y_2, \ldots, Y_n)$. On dit que les vecteurs aléatoires $\vec{X}$ et $\vec{Y}$ *ne sont pas corrélés* si aucune des composantes du vecteur $\vec{X}$ n'est corrélée avec aucune des composantes du vecteur $\vec{Y}$:

$$K_{x_i y_j} = M[\mathring{X}_i \mathring{Y}_j] = 0 \quad \text{pour} \quad i = 1, \ldots, n; \quad j = 1, \ldots, n.$$

CHAPITRE 9

LOI DE RÉPARTITION NORMALE D'UN SYSTÈME DE VARIABLES ALÉATOIRES

9.1. Loi normale dans le plan

De toutes les lois de répartition d'un couple de variables aléatoires la plus intéressante est la loi normale comme le plus fréquemment utilisée. Un système de deux variables aléatoires étant représenté par un point aléatoire dans le plan, la loi normale d'un couple est souvent appelée *loi normale dans le plan.*

Dans le cas général la densité de répartition conjointe de deux variables aléatoires normales s'exprime par la formule

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]}. \tag{9.1.1}$$

Cette loi dépend des cinq paramètres: m_x, m_y, σ_x, σ_y et r. Voyons le sens à attribuer à ces paramètres. Nous allons montrer que les paramètres m_x, m_y sont les espérances mathématiques (centres de dispersion) des variables X et Y; σ_x, σ_y leurs écarts quadratiques moyens et r le coefficient de corrélation de X et Y.

Pour s'en rendre compte nous allons tout d'abord trouver les densités de probabilité marginales des variables du système. En vertu de la formule (8.4.2):

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \times$$

$$\times \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]} dy.$$

Calculons l'intégrale

$$I = \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]} dy.$$

A cet effet, posons:

$$\frac{x-m_x}{\sigma_x\sqrt{2}}=u, \qquad \frac{y-m_y}{\sigma_y\sqrt{2}}=v; \tag{9.1.2}$$

On a alors:

$$I=\sigma_y\sqrt{2}\int_{-\infty}^{\infty} e^{-\frac{1}{1-r^2}[u^2-2ruv+v^2]}\,dv.$$

En calcul intégral on a:

$$\int_{-\infty}^{\infty} e^{-Ax^2\pm 2Bx+C}\,dx=\sqrt{\frac{\pi}{A}}\,e^{-\frac{AC-B^2}{A}}\ ^{*)}. \tag{9.1.3}$$

Dans notre cas:

$$A=\frac{1}{1-r^2}; \quad B=\frac{ru}{1-r^2}; \quad C=\frac{u^2}{1-r^2}.$$

En substituant ces valeurs dans la formule (9.1.3) on obtient:

$$I=\sigma_y\sqrt{2}\sqrt{\pi(1-r^2)}\,e^{-u^2},$$

d'où

$$f_1(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-u^2},$$

ou, compte tenu de (9.1.2):

$$f_1(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}. \tag{9.1.4}$$

Ainsi, la variable X suit une loi normale de centre de dispersion m_x et d'écart quadratique moyen σ_x.

D'une manière analogue on obtient:

$$f_2(y)=\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}, \tag{9.1.5}$$

c'est-à-dire que Y suit une loi normale de centre de dispersion m_y et d'écart quadratique moyen σ_y.

Il reste à démontrer que le paramètre r dans la formule (9.1.1) est le coefficient de corrélation des variables X et Y. Pour cela calculons la covariance:

$$K_{xy}=\int\int_{-\infty}^{\infty}(x-m_x)(y-m_y)f(x, y)\,dx\,dy,$$

où m_x, m_y sont les espérances mathématiques de X et Y.

*) Pour calculer l'intégrale (9.1.3) il suffit de compléter l'expression dans l'exposant de manière à en faire un carré entier et après le changement de variable d'utiliser l'intégrale d'Euler-Poisson (6.1.3).

En substituant dans cette formule l'expression de $f(x, y)$ on obtient:

$$K_{xy} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \int\limits_{-\infty}^{\infty}\int (x-m_x)(y-m_y)\, e^{-A(x,y)}\, dx\, dy, \qquad (9.1.6)$$

où

$$A(x, y) = -\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right].$$

On effectue dans l'intégrale double (9.1.6) un changement de variables en posant:

$$\frac{x-m_x}{\sigma_x\sqrt{2}} = u; \qquad \frac{1}{\sqrt{2(1-r^2)}}\left(\frac{y-m_y}{\sigma_y} - r\frac{x-m_x}{\sigma_x}\right) = w. \qquad (9.1.7)$$

Le jacobien de la transformation est:

$$2\sigma_x\sigma_y\sqrt{1-r^2},$$

par conséquent:

$$K_{xy} = \frac{1}{\pi}\int\limits_{-\infty}^{\infty}\int (u\sigma_x\sqrt{2})\,\sigma_y\sqrt{2(1-r^2)}\left(w + \frac{ru}{\sqrt{1-r^2}}\right) e^{-u^2-w^2}\, du\, dw =$$

$$= \frac{2\sigma_x\sigma_y\sqrt{1-r^2}}{\pi}\int\limits_{-\infty}^{\infty} u e^{-u^2}\, du \int\limits_{-\infty}^{\infty} w e^{-w^2}\, dw +$$

$$+ \frac{2\sigma_x\sigma_y r}{\pi}\int\limits_{-\infty}^{\infty} u^2 e^{-u^2}\, du \int\limits_{-\infty}^{\infty} e^{-w^2}\, dw.$$

Compte tenu de

$$\int\limits_{-\infty}^{\infty} u e^{-u^2}\, du = \int\limits_{-\infty}^{\infty} w e^{-w^2}\, dw = 0;$$

$$\int\limits_{-\infty}^{\infty} u^2 e^{-u^2}\, du = \frac{\sqrt{\pi}}{2}; \qquad \int\limits_{-\infty}^{\infty} e^{-w^2}\, dw = \sqrt{\pi},$$

on a:

$$K_{xy} = r\sigma_x\sigma_y; \qquad r = \frac{K_{xy}}{\sigma_x\sigma_y}. \qquad (9.1.8)$$

Nous avons ainsi montré que le paramètre r de la formule (9.1.1) est le coefficient de corrélation des variables X et Y.

Supposons maintenant que les variables aléatoires X et Y suivant une loi normale dans le plan ne soient pas corrélées; ceci conduit

à poser dans la formule (9.1.1) $r = 0$. Alors:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-m_x)^2}{2\sigma_x^2} - \frac{(y-m_y)^2}{2\sigma_y^2}} \quad (9.1.9)$$

Il est facile de voir que des variables (X, Y) normales de densité (9.1.9) sont non seulement non corrélées mais également *indépendantes*. En effet:

$$f(x, y) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} = f_1(x) f_2(y),$$

c'est-à-dire que la densité de probabilité du système est égale au produit des densités marginales, ce qui témoigne en faveur de l'indépendance des variables aléatoires (X, Y).

Ainsi, dans le cas d'un système de variables aléatoires normales la non-corrélation entraîne également leur indépendance. Dans le cas de la loi normale dire « variables non corrélées » c'est dire aussi « variables indépendantes ».

Pour $r \neq 0$ les variables aléatoires (X, Y) sont indépendantes. Il est facile de voir, en calculant les lois de répartition conditionnelles à l'aide des formules (8.4.6), que l'on a:

$$f(y \mid x) = \frac{1}{\sigma_y\sqrt{1-r^2}\sqrt{2\pi}} e^{-\frac{1}{2(1-r^2)}\left(\frac{y-m_y}{\sigma_y} - r\frac{x-m_x}{\sigma_x}\right)^2},$$

$$f(x \mid y) = \frac{1}{\sigma_x\sqrt{1-r^2}\sqrt{2\pi}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x-m_x}{\sigma_x} - r\frac{y-m_y}{\sigma_y}\right)^2}.$$

Etudions l'une de ces lois de répartition conditionnelles, par exemple $f(y \mid x)$. A cet effet nous allons écrire l'expression $f(y \mid x)$ comme suit:

$$f(y \mid x) = \frac{1}{\sigma_y\sqrt{1-r^2}\sqrt{2\pi}} e^{-\frac{1}{2(1-r^2)\sigma_y^2}\left[y-m_y-r\frac{\sigma_y}{\sigma_x}(x-m_x)\right]^2}.$$

Il est évident que c'est la densité de la loi normale de centre de dispersion

$$m_{y|x} = m_y + r\frac{\sigma_y}{\sigma_x}(x-m_x) \quad (9.1.10)$$

et d'écart quadratique moyen

$$\sigma_{y|x} = \sigma_y\sqrt{1-r^2}. \quad (9.1.11)$$

Les formules (9.1.10) et (9.1.11) montrent que dans la loi de répartition conditionnelle de Y seule l'espérance mathématique dépend de la valeur fixée de $X = x$, la variance en étant indépendante.

La grandeur $m_{y\mid x}$ est appelée espérance mathématique conditionnelle de Y pour x donné. La relation (9.1.10) peut être représentée dans le plan xOy en portant l'espérance mathématique conditionnelle $m_{y\mid x}$ en ordonnées. On obtient une droite appelée *droite de régression* de Y en X. D'une manière analogue, la droite d'équation

$$x = m_x + r\frac{\sigma_x}{\sigma_y}(y - m_y) \tag{9.1.12}$$

est la droite de régression de X en Y.

Les droites de régression coïncident seulement lorsque la relation fonctionnelle entre Y et X est linéaire. Quand X et Y sont indépendantes, les droites de régression sont parallèles aux axes de coordonnées.

En revenant à l'expression (9.1.1) de la densité de la répartition normale dans le plan on remarque que celle-ci est complètement déterminée par cinq paramètres: deux coordonnées m_x, m_y du centre de dispersion, deux écarts quadratiques moyens σ_x, σ_y et un coefficient de corrélation r. Les trois derniers paramètres sont à leur tour donnés par les éléments de la matrice de corrélation, à savoir par les variances D_x, D_y et la covariance K_{xy}. Ainsi, dans le cas de la loi normale, un nombre minimal de caractéristiques numériques du système, à savoir les espérances mathématiques, les variances et la covariance déterminent entièrement la loi de répartition, c'est-à-dire forment un système *exhaustif* de caractéristiques.

Etant donné que dans les applications pratiques on rencontre souvent la loi normale, la plupart du temps, pour caractériser entièrement la loi de répartition d'un système, il suffit d'en donner le nombre minimal — cinq — de caractéristiques numériques.

9.2. Ellipse de dispersion. Forme canonique de la loi normale

Considérons la surface de répartition représentant la fonction (9.1.1). Elle a la forme d'un monticule, dont le sommet se trouve audessus du point (m_x, m_y) (fig. 9.2.1).

Dans les sections de la surface de répartition par des plans parallèles à l'axe $f(x, y)$ on obtient des courbes semblables aux courbes de la répartition normale. Dans les sections de la surface de répartition par des plans parallèles au plan xOy on obtient des ellipses. L'équation de la projection d'une telle ellipse sur le plan xOy est:

$$\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} = \text{const},$$

ou en désignant la constante par λ^2,

$$\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} = \lambda^2. \tag{9.2.1}$$

L'équation (9.2.1) de l'ellipse peut être étudiée par les méthodes habituelles de la géométrie analytique. On montre aisément que le centre de l'ellipse (9.2.1) se trouve au point de coordonnées (m_x, m_y), les axes de symétrie de l'ellipse formant avec l'axe Ox des angles donnés par l'équation

$$\operatorname{tg} 2\alpha = \frac{2r\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \text{ *)}. \tag{9.2.2}$$

Cette équation donne deux valeurs de l'angle α : α_1 et α_2, différant de $\frac{\pi}{2}$.

Ainsi, l'orientation de l'ellipse (9.2.1) par rapport aux axes de coordonnées dépend directement du coefficient de corrélation r

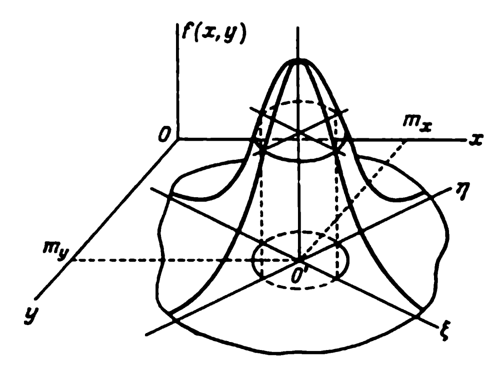


Fig. 9.2.1

du système (X, Y); si les variables X et Y ne sont pas corrélées (dans le cas considéré indépendantes), les axes de symétrie de l'ellipse sont parallèles aux axes de coordonnées.

En coupant la surface de répartition par les plans parallèles au plan xOy et en projetant les sections sur ce plan xOy on obtient une famille d'ellipses semblables, de même orientation et de centre commun (m_x, m_y). En tous les points d'une ellipse la densité de probabilité $f(x, y)$ est constante. C'est pour cela qu'on les appelle *ellipses d'égale densité* ou *ellipses de dispersion*. Les axes communs de toutes les ellipses sont les *axes principaux de dispersion*.

On sait que l'équation d'une ellipse prend une forme particulièrement simple, dite « canonique », lorsque les axes de coordonnées coïncident avec les axes de symétrie de l'ellipse. Pour rapporter l'équation de l'ellipse de dispersion à une forme canonique il suffit

*) Cf. formule (14.7.4).

de transférer l'origine des coordonnées au point (m_x, m_y) et de tourner les axes de coordonnées d'un angle donné par l'équation (9.2.2). La forme canonique de la loi normale dans le plan s'écrit :

$$f(\xi, \eta) = \frac{1}{2\pi\sigma_\xi\sigma_\eta} e^{-\frac{\xi^2}{2\sigma_\xi^2} - \frac{\eta^2}{2\sigma_\eta^2}}, \tag{9.2.3}$$

où σ_ξ, σ_η sont les écarts *quadratiques moyens principaux*, c'est-à-dire les écarts quadratiques moyens des variables aléatoires (Ξ, H) qui sont les coordonnées d'un point aléatoire dans le système des axes principaux de dispersion $O\xi$, $O\eta$. Les écarts quadratiques moyens principaux σ_ξ et σ_η s'expriment en fonction des écarts quadratiques moyens dans l'ancien système de coordonnées comme suit :

$$\left.\begin{aligned} \sigma_\xi^2 &= \sigma_x^2 \cos^2\alpha + r\sigma_x\sigma_y \sin 2\alpha + \sigma_y^2 \sin^2\alpha \\ \sigma_\eta^2 &= \sigma_x^2 \sin^2\alpha - r\sigma_x\sigma_y \sin 2\alpha + \sigma_y^2 \cos^2\alpha \end{aligned}\right\} ^{*}). \tag{9.2.4}$$

Généralement lorsqu'on envisage une loi de répartition normale dans le plan, on s'efforce d'orienter les axes de coordonnées Ox, Oy suivant les axes principaux de dispersion. Dans ce cas les écarts quadratiques moyens σ_x, σ_y seront les écarts quadratiques moyens principaux et la loi normale s'écrira :

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}}. \tag{9.2.5}$$

Dans certains cas on oriente les axes de coordonnées parallèlement aux axes de dispersion sans tout de même faire coïncider l'origine des coordonnées et le centre de dispersion. Les variables aléatoires (X, Y) se trouvent être indépendantes, l'expression de la loi normale étant

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-m_x)^2}{2\sigma_x^2} - \frac{(y-m_y)^2}{2\sigma_y^2}}, \tag{9.2.6}$$

m_x, m_y étant les coordonnées du centre de dispersion.

Passons dans la forme canonique de la loi normale (9.2.5) des écarts quadratiques moyens aux écarts probables :

$$E_x = \rho\sqrt{2}\sigma_x; \; E_y = \rho\sqrt{2}\sigma_y.$$

Les grandeurs E_x, E_y sont appelées *écarts probables principaux*. En exprimant σ_x, σ_y à l'aide de E_x, E_y dans l'équation (9.2.5) on

*) Pour la démonstration de ces formules voir § 14.7.

obtient une autre forme canonique de la loi normale:

$$f(x, y) = \frac{\rho^2}{\pi E_x E_y} e^{-\rho^2 \left(\frac{x^2}{E_x^2} + \frac{y^2}{E_y^2}\right)}. \quad (9.2.7)$$

Cette forme de la loi normale est souvent utilisée dans la théorie du tir.

Ecrivons l'équation de l'ellipse de dispersion sous forme canonique, soit:

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = k^2 \quad \text{ou} \quad \frac{x^2}{(k\sigma_x)^2} + \frac{y^2}{(k\sigma_y)^2} = 1, \quad (9.2.8)$$

où k est une constante.

Cette équation montre que les demi-axes de l'ellipse de dispersion sont proportionnels aux écarts quadratiques moyens principaux (donc également aux écarts probables principaux).

On appelle *ellipse de dispersion unitaire* celle des ellipses d'égale densité, dont les demi-axes sont égaux aux écarts quadratiques moyens σ_x, σ_y. (Si on utilise pour caractériser une dispersion non pas les écarts quadratiques moyens principaux mais les écarts probables principaux on appelle *ellipse unitaire* l'ellipse dont les demi-axes sont E_x, E_y).

D'une grande utilité est également la notion d'*ellipse de dispersion totale*. C'est une ellipse d'égale densité qui englobe toutes les valeurs possibles d'un couple de variables avec une certitude suffisamment élevée. Il est clair que les dimensions de cette ellipse dépendent de ce qu'on entend par « certitude suffisamment élevée ». En particulier, si on peut se contenter d'une probabilité de l'ordre de 0,99, l'ellipse de dispersion totale sera l'ellipse de demi-axes $3\sigma_x$, $3\sigma_y$.

Arrêtons-nous sur un cas particulier où les écarts quadratiques moyens principaux sont égaux entre eux:

$$\sigma_x = \sigma_y = \sigma.$$

Dans ce cas les ellipses de dispersion deviendront des cercles et la dispersion est dite *circulaire*. Ce cas présente cette particularité que chacun des axes passant par le centre de dispersion peut être considéré comme l'axe principal ou, en d'autres termes, la direction des axes principaux est indéterminée. Pour une dispersion non circulaire, des variables aléatoires X, Y réparties suivant une loi normale dans le plan sont indépendantes seulement si les axes de coordonnées sont parallèles aux axes principaux de dispersion et ceci quel que soit le choix du système de coordonnées rectangulaire. Cette particularité de la dispersion circulaire la rend bien plus commode que la dispersion elliptique. C'est pour cela qu'on essaie dans la mesure du possible de remplacer des dispersions non circulaires par des dispersions circulaires.

9.3. Probabilité de tomber dans un rectangle de côtés parallèles aux axes principaux de dispersion

Supposons que le point aléatoire (X, Y) suive la loi normale

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-m_x)^2}{2\sigma_x^2} - \frac{(y-m_y)^2}{2\sigma_y^2}}. \qquad (9.3.1)$$

Dans ce cas les axes principaux de dispersion sont parallèles aux axes de coordonnées et les variables X et Y sont indépendantes.

Il y a lieu de calculer la probabilité pour le point aléatoire (X, Y) de se trouver dans le rectangle R dont les côtés sont parallèles aux

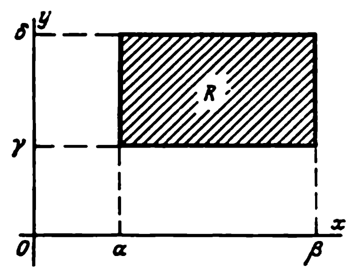


Fig. 9.3.1

axes de coordonnées Ox, Oy et par conséquent aux axes principaux de dispersion (fig. 9.3.1). En vertu de la formule (8.3.4) on a :

$$P((X, Y) \subset R) = \int_\alpha^\beta \int_\gamma^\delta f(x, y)\, dx\, dy =$$

$$= \int_\alpha^\beta \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} dx \int_\gamma^\delta \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} dy,$$

d'où, utilisant la formule (6.3.3) pour la probabilité de tomber dans un intervalle déterminé, on obtient :

$$P((X, Y) \subset R) = \left[\Phi^*\left(\frac{\beta-m_x}{\sigma_x}\right) - \Phi^*\left(\frac{\alpha-m_x}{\sigma_x}\right)\right] \times$$

$$\times \left[\Phi^*\left(\frac{\delta-m_y}{\sigma_y}\right) - \Phi^*\left(\frac{\gamma-m_y}{\sigma_y}\right)\right], \qquad (9.3.2)$$

où $\Phi^*(x)$ est la fonction de répartition normale.

Si la loi normale dans le plan est donnée sous forme canonique on a $m_x = m_y = 0$ et la formule (9.3.2) devient :

$$P((X, Y) \subset R) =$$

$$= \Phi^*\left(\frac{\beta}{\sigma_x}\right) - \Phi^*\left(\frac{\alpha}{\sigma_x}\right)\Big]\Big[\Phi^*\left(\frac{\delta}{\sigma_y}\right) - \Phi^*\left(\frac{\gamma}{\sigma_y}\right)\Big]. \qquad (9.3.3)$$

Si les côtés du rectangle ne sont pas parallèles aux axes de coordonnées, les formules (9.3.2) et (9.3.3) ne sont plus applicables. Ce n'est que dans le cas de la dispersion circulaire que la probabilité de tomber dans un rectangle d'orientation quelconque se calcule par l'une des formules (9.3.2) ou (9.3.3).

On utilise fréquemment les formules (9.3.2) et (9.3.3) lors des calculs des probabilités d'atteindre des cibles rectangulaires, presque rectangulaires, formées de rectangles ou susceptibles d'être approximativement remplacées par des rectangles.

Exemple. Un appareil est vérifié suivant deux paramètres X et Y. Ces deux paramètres sont indépendants et répartis suivant une loi normale de paramètres

$$m_x = 2, \quad \sigma_x = 0,8; \qquad m_y = 4, \quad \sigma_y = 0,5.$$

Pour que l'appareil soit accepté il faut et il suffit que le paramètre X se trouve

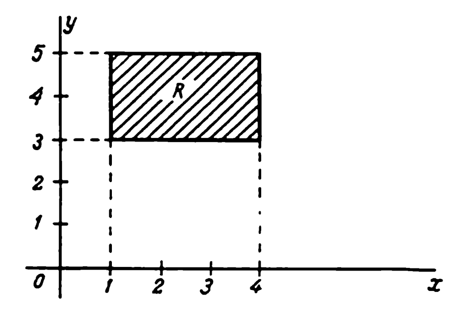


Fig. 9.3.2

entre 1 et 4, et le paramètre Y entre 3 et 5. Trouver la probabilité pour que l'appareil satisfasse à ces conditions.

Solution. Il y a lieu de trouver la probabilité pour le point aléatoire (X, Y) de se trouver dans le rectangle R représenté sur la figure 9.3.2.

En utilisant la table 1 de l'annexe on trouve:

$$P((X, Y) \subset R) = \left[\Phi^*\left(\frac{4-2}{0,8}\right) - \Phi^*\left(\frac{1-2}{0,8}\right)\right]\left[\Phi^*\left(\frac{5-4}{0,5}\right) - \Phi^*\left(\frac{3-4}{0,5}\right)\right] =$$
$$= [\Phi^*(2,5) - \Phi^*(-1,25)][\Phi^*(2) - \Phi^*(-2)] =$$
$$= [0,9938 - 0,1056][0,9772 - 0,0228] = 0,848.$$

9.4. Probabilité de tomber dans l'ellipse de dispersion

L'ellipse de dispersion (ellipse d'égale densité) est l'une des rares figures planes à l'intérieur desquelles un point peut se trouver avec une probabilité calculable sous forme explicite.

Supposons que l'on se donne la forme canonique de la loi normale dans le plan :

$$f(x, y) = \frac{1}{2\pi\sigma_y\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}. \tag{9.4.1}$$

Considérons l'ellipse de dispersion B_k d'équation

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = k^2,$$

où le paramètre k est le rapport des demi-axes de l'ellipse de dispersion aux écarts quadratiques moyens principaux. En vertu de la formule générale (8.3.3) on a :

$$P((X, Y) \subset B_k) = \iint_{(B_k)} f(x, y)\,dx\,dy =$$

$$= \iint_{(B_k)} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dx\,dy. \tag{9.4.2}$$

Après le changement de variables

$$\frac{x}{\sigma_x} = u; \qquad \frac{y}{\sigma_y} = v$$

l'ellipse B_k devient un cercle C_k de rayon k. Par conséquent :

$$P((X, Y) \subset B_k) = \frac{1}{2\pi} \iint_{(C_k)} e^{-\frac{u^2}{2} - \frac{v^2}{2}} du\,dv. \tag{9.4.3}$$

Passons dans l'intégrale (9.4.3) du système cartésien au système polaire, en posant :

$$u = r\cos\theta;\ v = r\sin\theta. \tag{9.4.4}$$

Le jacobien de la transformation (9.4.4) est égal à r. Après le changement de variables en question l'expression (9.4.3) devient :

$$P((X, Y) \subset B_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_0^k r e^{-\frac{r^2}{2}} dr\,d\theta = \int_0^k r e^{-\frac{r^2}{2}} dr = 1 - e^{-\frac{k^2}{2}}.$$

Ainsi, la probabilité pour un point aléatoire de se trouver à l'intérieur de l'éllipse de dispersion dont les demi-axes sont k fois les écarts quadratiques moyens est :

$$P((X, Y) \subset B_k) = 1 - e^{-\frac{k^2}{2}}. \tag{9.4.5}$$

A titre d'exemple, nous allons calculer la probabilité pour un point aléatoire réparti suivant une loi normale dans le plan xOy de

se trouver dans l'ellipse de dispersion unitaire dont les demi-axes sont

$$a = \sigma_x ; \; b = \sigma_y .$$

Pour une telle ellipse $k = 1$. On a alors

$$P((X, Y) \subset B_1) = 1 - e^{-\frac{1}{2}} .$$

Dans la table 2 de l'annexe on trouve

$$P((X, Y) \subset B_1) \approx 0{,}393.$$

La formule (9.4.5) est principalement utilisée pour calculer la probabilité de tomber dans un cercle dans le cas de la dispersion circulaire.

Exemple. Un nuage d'éclats représentant un disque plan (cercle) de rayon $R = 30$ m barre le chemin à un objectif mouvant de petite dimension, de surface 1,2 m. A l'intérieur du disque la densité des éclats est constante et égale à 2 éclats par m^2. Si le nuage recouvre entièrement l'objectif, le nombre d'éclats le touchant est réparti suivant une loi de Poisson. L'objectif étant petit, on peut le supposer ponctuel et considérer qu'ou bien il est entièrement recouvert par le nuage d'éclats (si son centre se trouve à l'intérieur du cercle) ou n'est pas du tout recouvert (si son centre ne se trouve pas dans le cercle). Le but est atteint lorsqu'au moins un éclat l'ait touché. Lors de la visée on s'efforce à faire coïncider dans le plan xOy le centre O_1 du cercle avec l'origine des coordonnées O (centre de l'objectif), mais les erreurs de visée font que le point O_1 se trouve dispersé au voisinage de O (fig. 9.4.1). La loi de dispersion est normale, la dispersion est circulaire avec $\sigma = 20$ cm. Trouver la probabilité $P(A)$ d'atteindre le but.

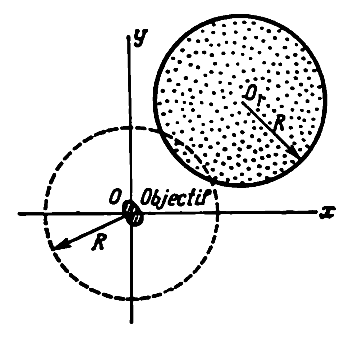


Fig. 9.4.1

Solution. Pour que le but soit atteint par les éclats, il faut que les deux événements suivants coïncident: 1) l'objectif (point O) entre dans le nuage d'éclats (cercle de rayon R) et 2) au moins un éclat le touche à condition que le premier événement se soit réalisé.

La probabilité de l'événement: « objectif entre dans le cercle » est de toute évidence égale à la probabilité pour que le centre du cercle (point aléatoire O_1) se trouve dans un cercle de rayon R avec le centre à l'origine des coordonnées. Appliquant la formule (9.4.5), on a:

$$k = \frac{30}{20} = 1{,}5.$$

La probabilité cherchée est:

$$p = 1 - e^{-\frac{1{,}5^2}{2}} \approx 0{,}675.$$

Puis nous allons trouver la probabilité p^* pour que l'objectif soit atteint à condition d'être recouvert par le disque d'éclats. Le nombre d'éclats atteignant l'objectif est égal au produit de l'aire de l'objectif par la densité du nuage

d'éclats:

$$n = 1{,}2 \cdot 2 = 2{,}4.$$

La probabilité conditionnelle p^* d'atteindre le but n'est rien d'autre que la probabilité qu'au moins un éclat l'atteigne. En utilisant la formule (5.9.5) du chapitre 5 on a :

$$p^* = R_1 = 1 - e^{-a} \approx 0{,}909.$$

La probabilité pour que l'objectif soit atteint est:

$$P(A) = 0{,}675 \cdot 0{,}909 \approx 0{,}613.$$

Nous nous servons de la formule (9.4.5) pour énoncer une loi fort importante pour les applications pratiques, à savoir la *loi de Rayleigh*.

Considérons un point aléatoire (X, Y) dans le plan xOy (fig. 9.4.2) se dispersant autour de l'origine des coordonnées O suivant une loi normale circulaire, d'écart quadratique moyen σ. Nous allons trouver la loi de répartition de la grandeur aléatoire R, distance du point (X, Y) à l'origine des coordonnées, c'est-à-dire de la longueur du vecteur aléatoire de composantes X, Y.

Cherchons tout d'abord la fonction de répartition $F(r)$ de R. Par définition

$$F(r) = P(R < r).$$

Ce n'est rien d'autre que la probabilité pour le point aléatoire (X, Y) de se trouver à l'intérieur d'un cercle de rayon r (fig. 9.4.2). La formule (9.4.5) permet de trouver cette probabilité:

$$F(r) = 1 - e^{-\frac{k^2}{2}};$$

ici $k = \frac{r}{\sigma}$, donc

$$F(r) = 1 - e^{-\frac{r^2}{2\sigma^2}}. \tag{9.4.6}$$

Cette expression de la fonction de répartition n'a de sens que pour des r positifs ; pour $r < 0$ on doit poser $F(r) = 0$.

En dérivant la fonction de répartition $F(r)$ par rapport à r on trouve la densité de probabilité:

$$f(r) = \begin{cases} \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} & \text{pour } r > 0, \\ 0 & \text{pour } r < 0. \end{cases} \tag{9.4.7}$$

La loi de Rayleigh (9.4.7) se rencontre dans de nombreuses applications pratiques; citons parmi d'autres la théorie du tir, la radiotechnique, l'électrotechnique, etc.

Le graphique de la fonction $f(r)$ (densité de la loi de Rayleigh) est donné sur la figure 9.4.3.

Calculons les caractéristiques numériques de la grandeur R répartie suivant la loi de Rayleigh : son mode $\mathcal{M}$ et son espérance mathématique m_r. Pour trouver le mode, qui est l'abscisse du point

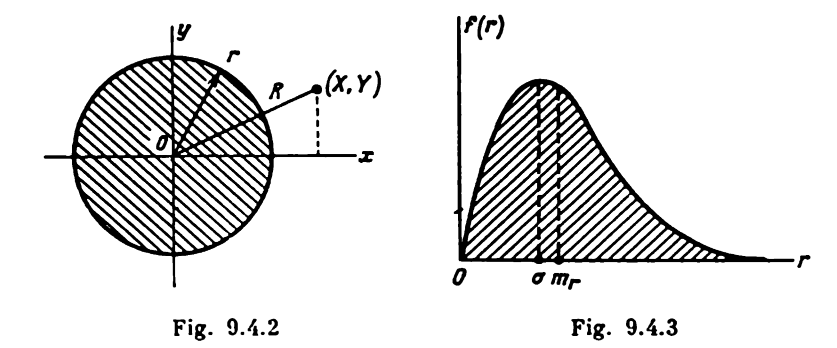


Fig. 9.4.2

Fig. 9.4.3

où la densité de probabilité est maximale, il y a lieu de dériver $f(r)$ et d'annuler cette dérivée, on obtient :

$$1-\frac{r^2}{\sigma^2}=0; \qquad \sigma^2=r^2.$$

La racine de cette équation est le mode cherché

$$\mathcal{M} = \sigma. \tag{9.4.8}$$

Ainsi, la valeur la plus probable de la distance R du point aléatoire (X, Y) à l'origine des coordonnées est égale à l'écart quadratique moyen de la dispersion.

L'espérance mathématique m_r est donnée par la formule :

$$m_r=\int_0^\infty rf(r)\,dr=\int_0^\infty \frac{r^2}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}\,dr.$$

Après le changement de variables

$$\frac{r}{\sigma\sqrt{2}}=t$$

on obtient :

$$m_r=\sigma\sqrt{2}\int_0^\infty 2t^2e^{-t^2}\,dt=\sigma\sqrt{2}\int_0^\infty t\cdot 2te^{-t^2}\,dt.$$

L'intégration par parties donne :

$$m_r=\sigma\sqrt{2}\,\frac{\sqrt{\pi}}{2}=\sigma\sqrt{\frac{\pi}{2}}\approx 1{,}25\sigma. \tag{9.4.9}$$

9.5. Probabilité de tomber dans un domaine de forme arbitraire

Dans le tir par obus percutants la probabilité d'atteindre le but est la probabilité pour le point (X, Y) de se trouver dans un certain domaine D. Supposons que le point aléatoire (X, Y) suive une loi normale et que cette loi soit donnée sous forme canonique. La probabilité pour le point (X, Y) d'atteindre le domaine D est donnée par l'intégrale :

$$P((X, Y) \subset D) = \frac{1}{2\pi\sigma_x\sigma_y} \iint\limits_{(D)} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}} \, dx \, dy. \qquad (9.5.1)$$

Dans certains cas particuliers (par exemple lorsque le domaine D est un rectangle dont les côtés sont parallèles aux axes principaux de dispersion ou si c'est une ellipse de dispersion) l'intégrale (9.5.1) peut être exprimée par des fonctions connues, ce qui n'est pas possible dans le cas général. Il est d'usage d'employer l'une des méthodes approchées suivantes.

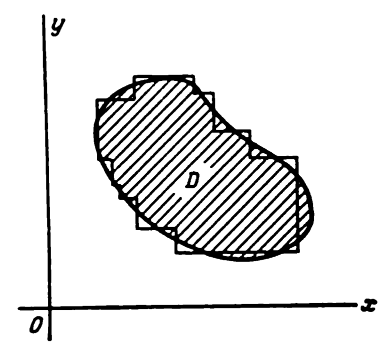


Fig. 9.5.1

1. Le domaine D est remplacé d'une manière approchée par un domaine formé de rectangles dont les côtés sont parallèles aux axes principaux de dispersion (fig. 9.5.1). La probabilité de tomber dans chacun de ces rectangles est donnée par la formule (9.3.3). On peut recommander cette méthode dans le cas où le nombre de rectangles remplaçant le domaine D n'est pas trop important.

2. Tout le plan xOy est divisé d'avance par un quadrillé (rectiligne ou courbe). Les probabilités d'atteindre chacune des cases peuvent s'exprimer exactement par des fonctions connues et l'on calcule ces probabilités. Ce quadrillage complété des probabilités d'atteindre les cases correspondantes est appelé *réseau de dispersion*. L'usage d'un tel réseau est simple : on applique le réseau sur l'image de l'objectif et on prend la somme des probabilités d'atteindre les cases recouvertes par l'objectif ; s'il recouvre une partie seulement de la case, la probabilité est prise proportionnellement a l'aire recouverte.

On peut utiliser le réseau de dispersion de deux manières différentes : a) dessiner le but à l'échelle du réseau, b) dessiner le réseau à l'échelle du but.

Lorsque la forme du but est compliquée et surtout si celui-ci n'est pas grand, il est plus commode de tracer sur l'image du but et à son échelle la partie du réseau occupée par celui-ci. Si au contrai-

re la forme du but est relativement simple et ses dimensions sont assez grandes (c'est-à-dire que le but occupe une partie importante de l'ellipse de dispersion), il est plus commode de dessiner le but à l'échelle du réseau. Comme le réseau normalisé correspond à une dispersion circulaire et les dispersions réelles ne le sont pas en général,

y

1	0	0	1	0	0	0	0	1	0	0	0	0	0	0					
1	0	1	0	1	0	0	1	0	0	0	0	0	0	1					
1	1	0	1	0	1	0	1	0	0	1	0	0	0	0					
2	1	1	1	1	1	0	1	0	1	0	0	1	0	0					
2	2	1	1	1	1	1	1	0	1	1	0	0	0	0					
2	2	2	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	0
2	2	2	2	2	2	2	1	1	1	1	1	1	1	0	0	0	0	0	0
3	3	2	2	2	2	2	2	1	1	1	1	1	1	0	0	1	0	0	0
4	3	3	3	3	3	2	2	2	2	1	1	1	1	1	0	0	0	0	0
5	4	4	4	3	3	3	3	2	2	2	1	1	1	1	1	0	1	0	0
5	5	5	5	4	4	4	3	2	2	2	2	1	1	1	1	1	0	0	0
6	6	6	5	5	5	4	4	3	2	2	2	1	1	1	0	0	0	0	1
7	7	6	6	6	5	5	4	4	3	2	2	2	1	1	1	1	1	1	0
8	8	7	7	6	6	6	5	4	4	3	2	2	2	1	1	0	0	0	0
9	9	8	8	7	7	6	5	5	4	3	3	2	2	1	1	1	1	0	0
10	9	9	8	8	7	6	6	5	4	3	3	2	2	1	1	1	0	1	0
10	10	10	10	8	8	7	5	5	5	4	3	2	2	1	2	1	1	0	1
11	11	10	10	9	8	7	6	6	5	4	3	2	2	2	1	1	0	1	0
11	11	11	10	9	9	8	7	6	5	4	3	3	2	2	2	1	1	0	0
12	11	11	10	10	9	8	7	6	5	5	4	3	2	2	2	2	1	1	1

x

Fig. 9.5.2

en traçant le but à l'échelle du réseau on est obligé d'utiliser des échelles différentes suivant les axes Ox et Oy. Pour faciliter l'opération il est commode de décalquer un réseau de dispersion sur un papier transparent que l'on superpose à l'image du but. Sur la figure 9.5.2 on a donné à titre d'exemple un réseau rectiligne de dispersion pour un quadrant. Le côté d'une case est égal à $0,2\,E \approx 0,133\ \sigma$. On a inscrit dans les cases les probabilités de les atteindre, exprimées en quarantièmes de pour cent.

3. Lorsque les dimensions du domaine D ne sont pas importantes par rapport aux écarts quadratiques moyens (sont en-dessous de 0,5 à 1σ dans les directions des axes de coordonnées), la probabilité d'atteindre ce domaine peut être calculée par une formule approchée ne contenant pas d'intégration.

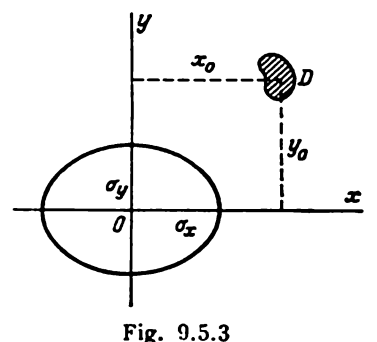


Fig. 9.5.3

Soit dans le plan xOy un petit but D de forme quelconque (fig. 9.5.3) dont les dimensions sont petites par rapport aux écarts probables E_x, E_y. En vertu de la formule générale on a :

$$P((X, Y) \subset D) = \iint_{(D)} f(x, y)\, dx\, dy, \tag{9.5.2}$$

où $f(x, y)$ est la densité de probabilité du système (X, Y). En appliquant à l'intégrale (9.5.2) le théorème de la moyenne, on a :

$$P((X, Y) \subset D) = f(x_0, y_0) \iint_{(D)} dx\, dy = f(x_0, y_0)\, S_D,$$

où (x_0, y_0) est un certain point à l'intérieur du domaine D; S_D l'aire du domaine D.

Dans le cas où la loi du système (X, Y) est normale et donnée sous forme canonique, on a :

$$P((X, Y) \subset D) = \frac{S_D}{2\pi\sigma_x\sigma_y} e^{-\frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2}}. \tag{9.5.3}$$

Les dimensions du domaine D n'étant pas importantes, la densité de probabilité $f(x, y)$ varie peu dans ce domaine et peut être supposée constante. Dans ce cas on peut prendre pour le point (x_0, y_0) un point quelconque du domaine D (par exemple, le centre approximatif du but).

Les formules du type (9.5.3) sont couramment utilisées dans les applications pratiques. Pour les domaines dont les dimensions maximales ne dépassent pas $0{,}5\sigma$ dans la direction correspondante, la précision des résultats obtenus est tout à fait satisfaisante. Dans certains cas elles peuvent également être utilisées pour les domaines plus importants (de l'ordre d'un écart quadratique moyen). A condition d'introduire certaines corrections (en majorant un peu les valeurs de σ_x, σ_y) le domaine d'application de cette formule peut être étendu à des figures de deux écarts quadratiques moyens de dimension.

9.6. Loi normale dans l'espace à trois dimensions. Forme générale de la loi normale pour un système d'un nombre arbitraire de variables aléatoires

Lors de l'étude des questions liées au tir par obus fusants on a affaire à la loi de répartition des points d'explosion de l'obus dans l'espace. Lorsque l'on utilise des fusées fusantes ordinaires, cette loi de répartition est approximativement normale.

Dans le présent paragraphe nous n'envisagerons que la forme canonique de la loi normale dans l'espace:

$$f(x, y, z) = \frac{1}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}+\frac{z^2}{\sigma_z^2}\right)}, \qquad (9.6.1)$$

où σ_x, σ_y, σ_z sont les écarts quadratiques moyens principaux.

En passant des écarts quadratiques moyens aux écarts probables on a:

$$f(x, y, z) = \frac{\rho^3}{\pi^{3/2} E_x E_y E_z} e^{-\rho^2\left(\frac{x^2}{E_x^2}+\frac{y^2}{E_y^2}+\frac{z^2}{E_z^2}\right)}. \qquad (9.6.2)$$

Dans les problèmes de tir par obus fusants, il y a parfois lieu de calculer la probabilité d'explosion d'un obus à l'intérieur d'un domaine D donné. Dans le cas général cette probabilité s'exprime par l'intégrale triple suivante:

$$P((X, Y, Z) \subset D) = \iiint_{(D)} f(x, y, z)\, dx\, dy\, dz. \qquad (9.6.3)$$

En général on ne peut exprimer l'intégrale (9.6.3) au moyen de fonctions élémentaires. Cependant, il existe un certain nombre de domaines pour lesquels il est facile de trouver les probabilités.

1. *Probabilité d'atteindre un parallélépipède rectangulaire de côtés parallèles aux axes principaux de dispersion*

Soit R un parallélépipède rectangulaire limité par les abscisses α, β, les ordonnées γ, δ et les cotes ε, η (fig. 9.6.1). La probabilité de tomber dans le domaine R est de toute évidence égale à;

$$P((X, Y, Z) \subset R) = \left[\Phi^*\left(\frac{\beta}{\sigma_x}\right) - \Phi^*\left(\frac{\alpha}{\sigma_x}\right)\right] \times$$
$$\times \left[\Phi^*\left(\frac{\delta}{\sigma_y}\right) - \Phi^*\left(\frac{\gamma}{\sigma_y}\right)\right]\left[\Phi^*\left(\frac{\eta}{\sigma_z}\right) - \Phi^*\left(\frac{\varepsilon}{\sigma_z}\right)\right]. \qquad (9.6.4)$$

2. *Probabilité d'atteindre un ellipsoïde d'égale densité*

Considérons l'ellipsoïde B_k d'égale densité d'équation

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2} = k^2.$$

Les demi-axes de cot ellipsoïde sont proportionnels aux écarts quadratiques moyens principaux:

$$a = k\sigma_x;\ b = k\sigma_y;\ c = k\sigma_z.$$

En utilisant la formule (9.6.1) pour $f\ (x,\ y,\ z)$ écrivons la probabilité d'atteindre l'ellipsoïde B_k:

$$P((X, Y, Z) \subset B_k) = \frac{1}{(2\pi)^{3/2}\,\sigma_x\sigma_y\sigma_z} \iiint\limits_{(B_k)} e^{-\frac{1}{2}\left(\frac{x}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)}\,dx\,dy\,dz.$$

Passons des coordonnées cartésiennes aux coordonnées polaires

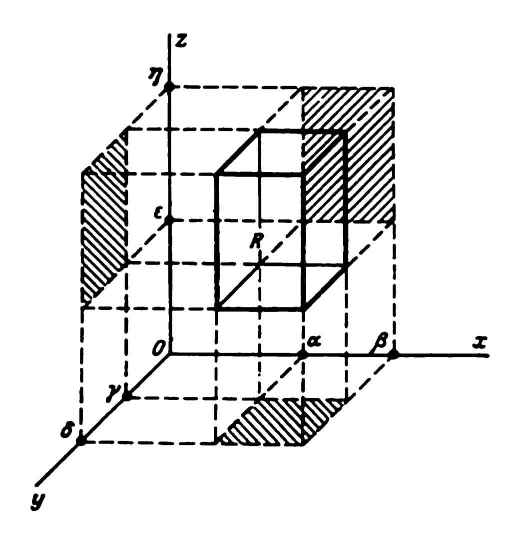


Fig. 9.6.1

(sphériques) par le changement de variables

$$\left.\begin{aligned} \frac{x}{\sigma_x} &= r\cos\theta\cos\varphi,\\ \frac{y}{\sigma_y} &= r\cos\theta\sin\varphi,\\ \frac{z}{\sigma_z} &= r\sin\theta. \end{aligned}\right\} \qquad (9.6.5)$$

Le jacobien de la transformation (9.6.5) est:

$$I = r^2\cos\theta\sigma_x\sigma_y\sigma_z.$$

En passant aux nouvelles variables, on a:

$$P((X, Y, Z) \subset B_k) =$$

$$= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_0^k \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} r^2 \cos\theta e^{-\frac{r^2}{2}} dr\, d\theta\, d\varphi = \sqrt{\frac{2}{\pi}} \int_0^k r^2 e^{-\frac{r^2}{2}} dr.$$

En intégrant par parties on obtient finalement:

$$P((X, Y, Z) \subset B_k) = \sqrt{\frac{2}{\pi}} \left\{ -ke^{-\frac{k^2}{2}} + \int_0^k e^{-\frac{r^2}{2}} dr \right\} =$$

$$= 2\Phi^*(k) - 1 - \sqrt{\frac{2}{\pi}} ke^{-\frac{k^2}{2}}. \qquad (9.6.6)$$

3. Probabilité d'atteindre un domaine cylindrique dont la génératrice est parallèle à l'un des axes principaux de dispersion

Soit un domaine cylindrique C dont la génératrice est parallèle à l'un des axes principaux de dispersion (par exemple à l'axe Oz) et la directrice est un contour du domaine quelconque D dans le plan xOy (fig. 9.6.2). Supposons que le cylindre C soit limité par

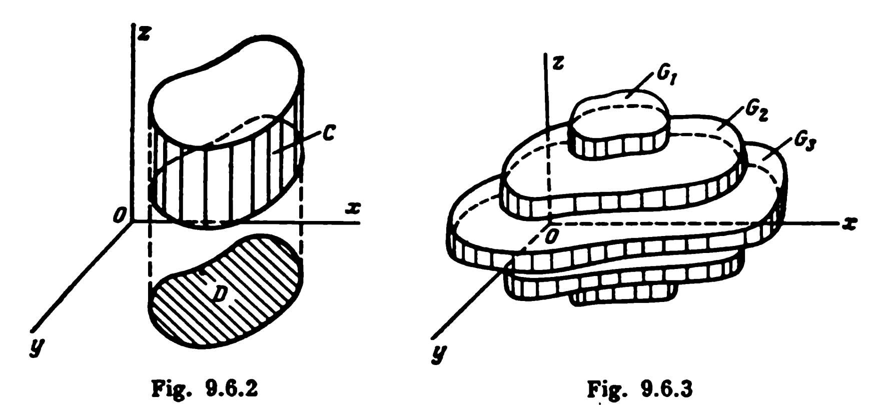


Fig. 9.6.2 Fig. 9.6.3

deux plans $z = \varepsilon$ et $z = \eta$. Calculons la probabilité d'atteindre le cylindre C. On conçoit aisément que c'est la probabilité de la réalisation simultanée de deux événements suivants: 1) le point (X, Y) tombe dans le domaine D, 2) la variable Z se trouve dans l'intervalle (ε, η). Les variables (X, Y, Z) suivant une loi normale sous forme canonique étant indépendantes, les deux événements en question le

sont également. Ainsi

$$P((X, Y, Z) \subset C) = P((X, Y) \subset D)\, P(\varepsilon < Z < \eta) =$$

$$= P((X, Y) \subset D)\left[\Phi^*\left(\frac{\eta}{\sigma_z}\right) - \Phi^*\left(\frac{\varepsilon}{\sigma_z}\right)\right]. \qquad (9.6.7)$$

La probabilité $P((X, Y) \subset D)$ figurant dans la formule (9.6.7) peut être calculée par l'une quelconque des méthodes de calcul des probabilités d'atteindre un domaine plan.

La formule (9.6.7) se trouve à la base de la méthode suivante de calcul des probabilités d'atteindre l'espace G de forme quelconque: on divise approximativement le domaine G en plusieurs domaines cylindriques G_1, G_2, . . . (fig. 9.6.3) et on calcule la probabilité d'atteindre chacun d'eux par la formule (9.6.7). Pour utiliser cette méthode il suffit de dessiner les figures qu'on aura dans les sections du domaine G par des plans parallèles à l'un des plans de coordonnées. La probabilité d'atteindre chacune de ces figures planes peut être calculée à l'aide du réseau de dispersion.

Pour conclure ce chapitre nous allons écrire l'expression générale de la loi normale dans l'espace pour un nombre quelconque de dimensions n.

La densité de probabilité de cette loi est:

$$f(x_1, x_2, \ldots, x_n) = \frac{\sqrt{|C|}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}(x_i - m_{x_i})(x_j - m_{x_j})}, \qquad (9.6.8)$$

où $|C|$ est le déterminant de la matrice $C = \|C_{ij}\|$, qui est la matrice inverse de la matrice de corrélation $K = \|K_{ij}\|$. On a:

$$C_{ij} = (-1)^{i+j} \frac{M_{ij}}{|K|},$$

où $|K|$ est le déterminant de la matrice de corrélation, et M_{ij} le mineur de ce déterminant obtenu en barrant la i-ème ligne et la j-ième colonne. Notons que

$$|C| = \frac{1}{|K|}.$$

De l'expression générale (9.6.8) s'obtiennent toutes les formes de la loi normale pour un nombre quelconque de dimensions et quelles que soient les relations entre les variables aléatoires. En

particulier pour $n = 2$ (dispersion dans le plan) la matrice de corrélation est

$$K = \begin{vmatrix} \sigma_x^2 & \sigma_x \sigma_y r \\ \sigma_x \sigma_y r & \sigma_y^2 \end{vmatrix},$$

où r est le coefficient de corrélation. D'où:

$$|K| = \sigma_x^2 \sigma_y^2 (1 - r^2); \quad |C| = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - r^2)};$$

$$C = \begin{vmatrix} \dfrac{1}{\sigma_x^2 (1 - r^2)} & \dfrac{-r}{\sigma_x \sigma_y (1 - r^2)} \\ \dfrac{-r}{\sigma_x \sigma_y (1 - r^2)} & \dfrac{1}{\sigma_y^2 (1 - r^2)} \end{vmatrix}.$$

En substituant le déterminant $|C|$ et les termes de la matrice C dans (9.6.8) on obtient la formule (9.1.1) pour la loi normale dans le plan par laquelle nous avons commencé le paragraphe 9.1.

CHAPITRE 10

CARACTÉRISTIQUES NUMÉRIQUES DES FONCTIONS DE VARIABLES ALÉATOIRES

10.1. Espérance mathématique. Variance

Pour résoudre différents problèmes liés aux phénomènes aléatoires la théorie moderne des probabilités fait appel aux lois des variables aléatoires, qu'il importe de connaître. On pourrait déterminer ces lois expérimentalement, mais en général les expériences visant à trouver la loi de répartition d'une variable ou d'un système de variables aléatoires (surtout dans le domaine militaire) sont compliquées et coûteuses. Le problème qui se pose tout naturellement est de réduire au minimum le volume des expériences et d'essayer de se faire une idée des lois inconnues en se basant sur les lois connues d'autres variables aléatoires. On procède par des méthodes indirectes, qui jouent un rôle particulièrement important dans la théorie des probabilités. En général la variable qui nous intéresse est une fonction d'autres variables aléatoires; connaissant la loi de répartition des arguments on arrive souvent à trouver la loi de la fonction. Plusieurs problèmes de ce genre seront étudiés au chap. 12.

Souvent dans la pratique on n'a pas besoin de déterminer complètement la loi de répartition d'une fonction des variables aléatoires, on peut se contenter de quelques caractéristiques numériques: il suffit de connaître l'espérance mathématique, la variance et parfois plusieurs moments d'ordre élevé. Il arrive souvent que les lois mêmes des arguments ne sont connues que très approximativement. Il s'agit alors de déterminer seulement les caractéristiques numériques des fonctions de variables aléatoires.

Examinons le problème suivant: soit Y une fonction de plusieurs variables aléatoires $X_1, X_2, \ldots, X_n$:

$$Y = \varphi(X_1, X_2, \ldots, X_n).$$

Supposant connue la loi de répartition du système d'arguments $(X_1, X_2, \ldots X_n)$, trouver les caractéristiques numériques de Y et en premier lieu, son espérance mathématique et sa variance.

Si l'on a pu déterminer la loi de répartition $g(y)$ de Y, il est facile de trouver les caractéristiques numériques; celles-ci sont

données par les formules:

$$m_y = \int_{-\infty}^{\infty} y g(y)\, dy;$$

$$D_y = \int_{-\infty}^{\infty} (y - m_y)^2 g(y)\, dy,$$

etc.

Cependant il est souvent difficile de trouver la loi de répartition $g(y)$ de Y. Mais pour résoudre le problème posé il n'est pas nécessaire de connaître cette loi, il suffit de savoir la loi des arguments $(X_1, X_2, \ldots, X_n)$. Dans certains cas pour trouver les caractéristiques numériques d'une fonction aléatoire il n'est même pas nécessaire de connaître la loi de répartition des arguments, les caractéristiques numériques des arguments suffisent.

On voit alors apparaître le problème de la détermination des caractéristiques numériques des fonctions des variables aléatoires sans en connaître les lois de répartition.

Commençons par déterminer les caractéristiques numériques d'une fonction, la loi de répartition des arguments étant donnée. Considérons le cas le plus simple, à savoir celui d'un seul argument.

Soit une variable aléatoire X de loi de répartition donnée et soit une autre variable aléatoire Y liée à X par la relation fonctionnelle:

$$Y = \varphi(X).$$

Calculer, sans trouver la loi de répartition de Y, l'espérance mathématique

$$m_y = M[\varphi(X)]. \tag{10.1.1}$$

Considérons tout d'abord le cas où X est une variable aléatoire discrète donnée par le tableau de répartition ci-dessous:

x_i	x_1	x_2	$\ldots$	x_n
p_i	p_1	p_2	$\ldots$	p_n

Nous allons écrire comme suit les valeurs possibles de Y et les probabilités correspondantes:

$\varphi(x_i)$	$\varphi(x_1)$	$\varphi(x_2)$	$\ldots$	$\varphi(x_n)$
p_i	p_1	p_2	$\ldots$	p_n

. (10.1.2)

Le tableau (10.1.2) n'est pas, en toute rigueur, celui de la variable Y, car dans le cas général certaines des valeurs

$$\varphi(x_1),\ \varphi(x_2),\ \ldots,\ \varphi(x_n) \tag{10.1.3}$$

peuvent coïncider; de plus les valeurs dans la ligne supérieure ne sont pas obligatoirement disposées par ordre croissant. Pour passer

du tableau (10.1.2) à la suite réelle de répartition de Y il aurait fallu disposer les valeurs (10.1.3) par ordre croissant et réunir les colonnes correspondant aux valeurs égales de Y en additionnant leurs probabilités. Cependant ici la vraie loi de répartition de Y ne nous intéresse pas; pour trouver l'espérance mathématique le tableau non ordonné (10.1.2) nous suffit. L'espérance mathématique de Y est donnée par la formule:

$$M[\varphi(X)] = \sum_i \varphi(x_i)\, p_i. \qquad (10.1.4)$$

Il est évident que la grandeur $m_y = M[\varphi(X)]$ ne peut changer si dans la somme on réunit d'avance certains termes ou on en change l'ordre.

Dans la formule (10.1.4) de l'espérance mathématique d'une fonction, la loi de répartition de la fonction n'apparaît pas d'une manière explicite, on y voit seulement la loi de répartition de l'argument. Ainsi, *pour trouver l'espérance mathématique d'une fonction il n'est pas nécessaire de connaître la loi de répartition de cette fonction, il suffit de connaître la loi de l'argument.*

En remplaçant dans la formule (10.1.4) la somme par une intégrale et la probabilité p_i par un élément de probabilité, on obtient une formule analogue pour une variable aléatoire continue:

$$M[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx, \qquad (10.1.5)$$

où $f(x)$ est la densité de probabilité de X.

De même on trouve l'espérance mathématique de la fonction $\varphi(X, Y)$ de deux arguments aléatoires X et Y. Pour des variables discrètes on a:

$$M[\varphi(X, Y)] = \sum_i \sum_j \varphi(x_i, y_j)\, p_{ij}, \qquad (10.1.6)$$

$p_{ij} = P((X = x_i)(Y = y_j))$ étant la probabilité pour que le système (X, Y) prenne la valeur (x_i, y_j).

Pour les variables continues on a:

$$M[\varphi(X, Y)] = \int\int_{-\infty}^{\infty} \varphi(x, y) f(x, y)\, dx\, dy, \qquad (10.1.7)$$

$f(x, y)$ étant la densité de probabilité du système (X, Y).

D'une manière analogue on peut trouver l'espérance mathématique d'une fonction d'un nombre quelconque d'arguments aléatoires.

Nous allons donner la formule dans le cas des variables continues:

$$M[\varphi(X_1, X_2, \ldots, X_n)] =$$
$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n, \quad (10.1.8)$$

$f(x_1, x_2, \ldots, x_n)$ étant la densité de probabilité du système $(X_1, X_2, \ldots, X_n)$.

Ainsi, l'espérance mathématique d'une fonction d'un nombre quelconque d'arguments aléatoires peut être calculée indépendamment de la loi de répartition de la fonction. D'une manière analogue on peut trouver d'autres caractéristiques numériques d'une fonction, à savoir ses moments d'ordre quelconque. Chaque moment étant l'espérance mathématique d'une certaine fonction de la variable aléatoire étudiée, on peut les calculer par des méthodes analogues à celles exposées ci-dessus. Nous ne donnerons ici que les formules de la variance et ceci pour le cas des arguments aléatoires continus.

La variance d'une fonction d'un argument aléatoire est donnée par la formule:

$$D[\varphi(X)] = \int_{-\infty}^{\infty} [\varphi(x) - m_\varphi]^2 f(x)\, dx, \quad (10.1.9)$$

où $m_\varphi = M[\varphi(x)]$ est l'espérance mathématique de la fonction $\varphi(X)$ et $f(x)$ la densité de probabilité de X.

La variance d'une fonction de deux arguments s'exprime par une formule analogue, soit:

$$D[\varphi(X, Y)] = \iint_{-\infty}^{\infty} [\varphi(x, y) - m_\varphi]^2 f(x, y)\, dx\, dy; \quad (10.1.10)$$

ici m_φ est l'espérance mathématique de la fonction $\varphi(X, Y)$ et $f(x, y)$ la densité de probabilité du système (X, Y).

Enfin dans le cas d'un nombre quelconque d'arguments, on a dans les mêmes notations:

$$D[\varphi(X_1, X_2, \ldots, X_n)] =$$
$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} [\varphi(x_1, x_2, \ldots, x_n) - m_\varphi]^2 f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n.$$
$$(10.1.11)$$

Remarquons que souvent dans les calculs de la variance il est commode d'utiliser la relation existant entre les moments initial

et centré du second ordre (voir chapitre 5) et d'écrire:

$$D[\varphi(X)] = \int_{-\infty}^{\infty} [\varphi(x)]^2 f(x)\,dx - m_\varphi^2; \qquad (10.1.12)$$

$$D[\varphi(X, Y)] = \iint_{-\infty}^{\infty} [\varphi(x, y)]^2 f(x, y)\,dx\,dy - m_\varphi^2; \qquad (10.1.13)$$

$$D[\varphi(X_1, X_2, \ldots, X_n)] =$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\varphi(x_1, x_2, \ldots, x_n)]^2 f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2 \ldots dx_n - m_\varphi^2. \qquad (10.1.14)$$

Les formules (10.1.12)-(10.1.14) peuvent être recommandées dans les cas où elles ne conduisent pas à des différences de nombres voisins, ce qui a lieu lorsque m_φ est assez petit.

Nous allons étudier quelques exemples pour utiliser les méthodes exposées ci-dessus.

Exemple 1. Soit dans le plan un segment de longueur l (fig. 10.1.1) tournant d'une manière aléatoire de telle sorte que toutes les directions sont

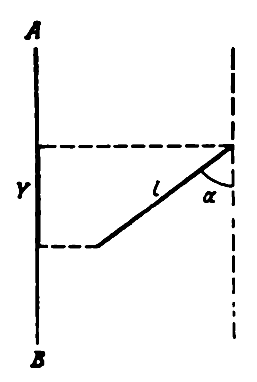


Fig. 10.1.1

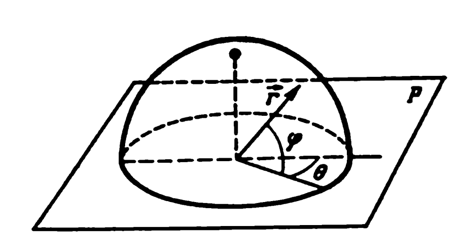


Fig. 10.1.2

équiprobables. Le segment est projeté sur un axe immobile AB. Trouver la valeur moyenne de la longueur de la projection du segment.

Solution. La longueur de la projection est égale à:

$$Y = l\,|\cos\alpha|.$$

où l'angle α est une variable aléatoire uniformément répartie sur l'intervalle 0, 2π.

La formule (10.1.5) donne:

$$m = M[l\,|\cos\alpha|] = \int_0^{2\pi} l\,|\cos\alpha|\,\frac{d\alpha}{2\pi} = \frac{2l}{\pi}\int_0^{\frac{\pi}{2}} \cos\alpha\,d\alpha = \frac{2l}{\pi} \approx 0{,}637l.$$

Exemple 2. Une barre, pouvant être schématisée par un segment de longueur l, vole en tournant autour de son centre de masse de telle sorte que toutes ses orientations dans l'espace sont équiprobables. Sur son chemin la barre rencontre un écran plan perpendiculaire à la direction de son mouvement et y laisse un trou. Trouver l'espérance mathématique de la longueur de ce trou.

Solution. Nous allons tout d'abord donner la forme mathématique à l'assertion « toutes les orientations d'une barre dans l'espace sont équiprobables ». La direction du segment l sera caractérisée par le vecteur unitaire $\vec{r}$ (fig. 10.1.2), dont l'orientation dans le système de coordonnées sphériques lié au plan P de projection est donnée par deux angles: θ dans le plan P et φ dans le plan perpendiculaire au plan P. Toutes les orientations du vecteur $\vec{r}$ étant équiprobables, toutes les positions de son extrémité sur la surface de la sphère S de rayon unité doivent avoir une même densité de probabilité; par conséquent l'élément de probabilité

$$f(\theta, \varphi)\, d\theta d\varphi,$$

où $f(\theta, \varphi)$ est la densité de probabilité conjointe des angles θ, φ, doit être proportionnel à la surface élémentaire ds sur la sphère S égale à:

$$ds = d\theta d\varphi \cos\varphi,$$

d'où:

$$f(\theta, \varphi)\, d\theta d\varphi = A \cos\varphi d\theta d\varphi; \quad f(\theta, \varphi) = A \cos\varphi,$$

A étant un coefficient de proportionnalité.

La valeur du coefficient A se trouve à partir de la relation:

$$\int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(\theta, \varphi)\, d\theta\, d\varphi = 1,$$

d'où

$$A = \frac{1}{4\pi}.$$

Ainsi, la densité de probabilité des angles θ, φ est donnée par la formule:

$$f(\theta, \varphi) = \frac{1}{4\pi} \cos\varphi \quad \text{pour} \begin{cases} 0 < \theta < 2\pi, \\ -\frac{\pi}{2} < \varphi < \frac{\pi}{2}. \end{cases} \tag{10.1.15}$$

Projetons le segment l sur le plan P, la longueur de la projection Y est égale à:

$$Y = l \cos\varphi.$$

Considérant Y comme une fonction des deux arguments θ et φ, on obtient à partir de la formule (10.1.7):

$$m_y = M[l \cos\varphi] = \frac{l}{4\pi} \int_0^{2\pi} d\theta \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2\varphi\, d\varphi = \frac{l}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2\varphi\, d\varphi = \frac{l\pi}{4} \approx 0{,}785 l.$$

Ainsi, la longueur moyenne du trou laissé par la barre dans l'écran est égale à 0,785 de sa longueur.

Exemple 3. Une figure plane d'aire S tourne d'une manière désordonnée dans l'espace de telle sorte que toutes ses orientations sont équiprobables.

Trouver l'aire moyenne de la projection de la figure S sur un plan immobile P (fig. 10.1.3).

S o l u t i o n. Convenons de caractériser l'orientation de la figure S dans l'espace par la direction de sa normale $\vec{N}$. Tout comme dans le problème précédent on utilisera un système de coordonnées sphériques lié au plan P. La direction de la normale $\vec{N}$ à la figure S sera caractérisée par les angles aléatoires θ et φ répartis avec la densité (10.1.5). L'aire Z de la projection de la figure S sur le plan P est égale à:

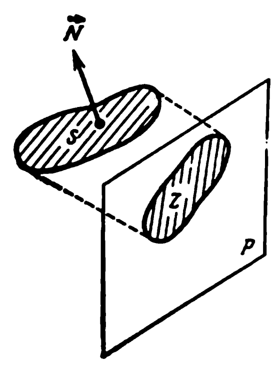


Fig. 10.1.3

$$Z=S\left|\cos\left(\frac{\pi}{2}-\varphi\right)\right|=S\,|\sin\varphi|$$

et l'aire moyenne de la projection est:

$$m_z=M\,[S\,|\sin\varphi|\,]=$$

$$=\frac{S}{4\pi}\int_0^{4\pi}d\theta\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos\varphi\,|\sin\varphi|\,d\varphi=\frac{S}{2}.$$

Ainsi, l'aire moyenne de la projection d'une figure plane orientée d'une façon quelconque dans l'espace est égale à la moitié de la surface de cette figure.

E x e m p l e 4. Lors de la poursuite d'un certain objectif par un radar, l'écho représentant cet objectif reste tout le temps sur l'écran. L'écran est un cercle C de rayon R. L'écho occupe sur l'écran une position aléatoire avec une densité de probabilité constante. Trouver la distance moyenne de l'écho au centre de l'écran.

S o l u t i o n. Désignant par D la distance cherchée on a $D=\sqrt{X^2+Y^2}$, où X, Y sont les coordonnées de l'écho; de plus on a $f(x, y)=\frac{1}{\pi R^2}$ dans les limites du cercle C et $f(x, y)=0$ à l'extérieur de ce cercle. En utilisant la formule (10.1.7) et en passant dans l'intégrale aux coordonnées polaires, on obtient:

$$m_D=\frac{1}{\pi R^2}\iint\limits_{(K)}\sqrt{x^2+y^2}\,dx\,dy=\frac{1}{\pi R^2}\int_0^{2\pi}d\varphi\int_0^R r^2\,dr=\frac{2}{3}R.$$

E x e m p l e 5. La fiabilité (probabilité de fonctionnement sans défaillance) d'un dispositif technique (d'un régulateur) est une certaine fonction $p(X, Y, Z)$ de trois paramètres. Les paramètres X, Y, Z sont des variables aléatoires de densité de probabilité conjointe donnée $f(x, y, z)$. Trouver la valeur moyenne (espérance mathématique) de la fiabilité du dispositif et l'écart quadratique moyen caractérisant sa stabilité.

S o l u t i o n. La fiabilité du dispositif $p(X, Y, Z)$ étant une fonction de trois variables aléatoires (paramètres) X, Y, Z, sa valeur moyenne (espérance mathématique) sera donnée par la formule (10.1.8):

$$m_p=M\,[p(X, Y, Z)]=\iiint\limits_{-\infty}^{\infty}p(x, y, z)\,f(x, y, z)\,dx\,dy\,dz. \quad (10.1.16)$$

La formule (10.1.14) donne:

$$D_p = D[p(X, Y, Z)] = \int\int\int_{-\infty}^{\infty} [p(x, y, z)]^2 f(x, y, z)\, dx\, dy\, dz - m_p^2,$$

$$\sigma = \sqrt{D_p}.$$

La formule (10.1.16) exprimant la probabilité moyenne (totale) de fonctionnement sans défaillance d'un dispositif compte tenu des grandeurs aléatoires dont dépend cette probabilité dans chaque cas, est un cas particulier de *la formule intégrale de la probabilité totale* généralisant la formule de la probabilité totale au cas d'un nombre infini (indénombrable) d'hypothèses.

Nous allons démontrer cette formule sous une forme générale.

Supposons que l'expérience dans laquelle apparaît l'événement A nous intéressant ait lieu dans des conditions aléatoires inconnues d'avance. Supposons que ces conditions soient caractérisées par des variables aléatoires continues

$$X_1, X_2, \ldots, X_n \qquad (10.1.17)$$

dont la densité de probabilité est:

$$f(x_1, x_2, \ldots, x_n).$$

La probabilité P_A d'apparition de l'événement A est une certaine fonction des variables aléatoires (10.1.17):

$$P_A(X_1, X_2, \ldots, X_n). \qquad (10.1.18)$$

Il y a lieu de trouver la valeur moyenne de cette probabilité ou en d'autres termes la probabilité de l'événement A:

$$P_A = M[P_A(X_1, X_2, \ldots, X_n)].$$

En utilisant la formule (10.1.8) donnant l'espérance mathématique d'une fonction on a:

$$P_A = \int\int_{-\infty}^{\infty} \ldots \int P_A(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n. \qquad (10.1.19)$$

La formule (10.1.19) porte le nom de formule intégrale de la probabilité totale. Il est facile de voir que de par sa structure elle est analogue à celle de la probabilité totale; en effet, il suffit de remplacer la suite discrète d'hypothèses par une gamme continue, la somme par une intégrale, la probabilité d'une hypothèse par un élément de probabilité

$$f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2, \ldots, dx_n,$$

et la probabilité conditionnelle d'un événement pour une hypothèse donnée par la probabilité conditionnelle de l'événement pour des

valeurs fixes des grandeurs aléatoires:

$$P_A(x_1, x_2, \ldots, x_n).$$

Souvent on utilise également la *formule intégrale de l'espérance mathématique totale*. Cette formule exprime l'espérance mathématique moyenne (totale) de la valeur qui sera prise par la variable aléatoire Z dans l'expérience dont les conditions sont inconnues d'avance (aléatoires). Si ces conditions sont caractérisées par des variables aléatoires continues

$$X_1, X_2, \ldots, X_n$$

de densité de probabilité

$$f(x_1, x_2, \ldots, x_n),$$

l'espérance mathématique de Z est une fonction des variables X_1, $X_2, \ldots, X_n$:

$$m_z(X_1, X_2, \ldots, X_n),$$

alors l'espérance mathématique totale de Z est donnée par la formule:

$$m_z = \int\int_{-\infty}^{\infty} \ldots \int m_z(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n, \tag{10.1.20}$$

appelée *formule intégrale de l'espérance mathématique totale*.

Exemple 6. L'espérance mathématique de la distance D où l'objectif sera détecté par quatre stations radar dépend de certains paramètres techniques de ces stations, soit:

$$X_1, X_2, X_3, X_4,$$

qui sont des variables aléatoires indépendantes de densité de probabilité

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4).$$

Pour des valeurs données des paramètres $X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3$, $X_4 = x_4$, l'espérance mathématique de la distance de détection est égale à:

$$m_D(x_1, x_2, x_3, x_4).$$

Trouver l'espérance mathématique moyenne (totale) de la distance de détection.

Solution. La formule (10.1.20) donne:

$$m_D = \int\int\int_{-\infty}^{\infty}\int m_D(x_1, x_2, x_3, x_4) f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4)\, dx_1\, dx_2\, dx_3\, dx_4.$$

10.2. Théorèmes sur les caractéristiques numériques

Dans le paragraphe précédent nous avons établi les formules permettant de trouver les caractéristiques numériques des fonctions lorsque les lois de répartition des arguments sont données. Cependant, souvent pour trouver les caractéristiques numériques des fonctions

il n'est pas indispensable de connaître les lois de répartition des arguments, il suffit d'en connaître certaines caractéristiques numériques ; dans ce cas on peut ignorer la loi de répartition. Le calcul des caractéristiques numériques des fonctions d'après celles des arguments est fréquemment utilisé en théorie des probabilités, simplifiant beaucoup la solution de nombreux problèmes. Ces méthodes sont principalement employées dans le cas des fonctions linéaires ; cependant certaines fonctions élémentaires non linéaires peuvent également être traitées.

Dans le présent paragraphe nous allons exposer des théorèmes sur les caractéristiques numériques des fonctions, qui, dans leur ensemble, forment un appareil assez simple de calcul des caractéristiques ayant un large champ d'application.

1. Espérance mathématique d'une variable non aléatoire

Si c est une variable non aléatoire on a :

$$M[c] = c.$$

Cette propriété est intuitivement évidente ; pour la démontrer on peut considérer la variable non aléatoire c comme un cas particulier des variables aléatoires, prenant sa seule valeur possible avec la probabilité unité :

$$M[c] = c \cdot 1 = c.$$

2. Variance d'une variable non aléatoire

Si c est une variable non aléatoire on a :

$$D[c] = 0.$$

D é m o n s t r a t i o n. Par définition de la variance :

$$D[c] = M[\overset{\circ}{c}{}^2] = M[(c - m_c)^2] = M[(c - c)^2] = M(0) = 0.$$

3. Mise en facteur d'une variable non aléatoire devant le signe de l'espérance mathématique

Si c est une variable non aléatoire et X une variable aléatoire on a :

$$M[cX] = cM[X], \qquad (10.2.1)$$

c'est-à-dire qu'*une variable non aléatoire peut être mise en facteur devant le signe de l'espérance mathématique.*

D é m o n s t r a t i o n.

a) Pour les variables discrètes:

$$M[cX] = \sum_i cx_i p_i = c \sum_i x_i p_i = cM[X].$$

b) Pour les variables continues:

$$M[cX] = \int_{-\infty}^{\infty} cxf(x)\,dx = c \int_{-\infty}^{\infty} xf(x)\,dx = cM[X].$$

4. Mise en facteur d'une grandeur non aléatoire devant le signe de la variance et de l'écart quadratique moyen

Si c est une variable non aléatoire et X une variable aléatoire on a:

$$D[cX] = c^2 D[X], \tag{10.2.2}$$

c'est-à-dire qu'*une variable non aléatoire peut être mise en facteur devant le signe de la variance à condition de l'élever au carré.*

D é m o n s t r a t i o n. Par définition de la variance on a:

$$D[cX] = M[(cX - M[cX])^2] =$$
$$= M[(cX - cm_x)^2] = c^2 M[(X - m_x)^2] = c^2 D[X].$$

C o r o l l a i r e

$$\sigma[cX] = |c|\,\sigma[X],$$

c'est-à-dire qu'on peut mettre en facteur devant le signe de l'écart quadratique moyen une variable non aléatoire en valeur absolue. Pour la démonstration il y a lieu d'extraire la racine carrée de l'expression (10.2.2), tenant compte du fait que l'écart quadratique moyen est une grandeur essentiellement positive.

5. Espérance mathématique d'une somme de variables aléatoires

Nous allons montrer que pour deux variables aléatoires quelconques X et Y on a:

$$M[X + Y] = M[X] + M[Y], \tag{10.2.3}$$

c'est-à-dire que l'*espérance mathématique d'une somme de deux variables aléatoires est égale à la somme de leurs espérances mathématiques.*

Cette propriété est connue sous le nom de *théorème d'addition des espérances mathématiques.*

D é m o n s t r a t i o n.

a) Soit (X, Y) un système de variables aléatoires discrètes. Nous allons appliquer à leur somme la formule générale (10.1.6) donnant

l'espérance mathématique d'une fonction de deux arguments:

$$M[X+Y]=\sum_i \sum_j (x_i+y_j)\,p_{ij}=\sum_i \sum_j x_i p_{ij}+\sum_i \sum_j y_j p_{ij}=$$
$$=\sum_i x_i \sum_j p_{ij}+\sum_j y_j \sum_i p_{ij}.$$

Mais $\sum_j p_{ij}$ n'est rien d'autre que la probabilité totale pour X de prendre la valeur x_i:

$$\sum_j p_{ij}=P(X=x_i)=p_i;$$

par conséquent:

$$\sum_i x_i \sum_j p_{ij}=\sum_i x_i p_i=M[X].$$

D'une manière analogue on peut montrer que

$$\sum_j y_j \sum_i p_{ij}=M[Y],$$

ce qui achève la démonstration.

b) Soit (X, Y) un système de variables aléatoires continues. En vertu de la formule (10.1.7) on a:

$$M[X+Y]=\int\limits_{-\infty}^{\infty}\int (x+y)\,f(x, y)\,dx\,dy=$$
$$=\int\limits_{-\infty}^{\infty}\int xf(x, y)\,dx\,dy+\int\limits_{-\infty}^{\infty}\int yf(x, y)\,dx\,dy. \qquad (10.2.4)$$

La première des intégrales (10.2.4) peut s'écrire comme suit:

$$\int\limits_{-\infty}^{\infty}\int xf(x, y)\,dx\,dy=\int\limits_{-\infty}^{\infty} x\left\{\int\limits_{-\infty}^{\infty} f(x, y)\,dy\right\}dx=\int\limits_{-\infty}^{\infty} xf_1(x)\,dx=M[X];$$

d'une manière analogue on a:

$$\int\limits_{-\infty}^{\infty}\int yf(x, y)\,dx\,dy=M[Y],$$

le théorème se trouve ainsi démontré.

Il y a lieu de noter spécialement que le théorème d'addition des espérances mathématiques est indifférent à la corrélation des variables aléatoires, étant vrai tant pour des variables indépendantes que liées.

Ce théorème peut être généralisé à un nombre quelconque de termes :

$$M\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} M[X_i], \tag{10.2.5}$$

c'est-à-dire que *l'espérance mathématique de la somme de plusieurs variables aléatoires est égale à la somme de leurs espérances mathématiques.*

Pour la démonstration il suffit de procéder par récurrence.

6. *Espérance mathématique d'une fonction linéaire*

Considérons une fonction linéaire de plusieurs arguments aléatoires $X_1, X_2, \ldots, X_n$:

$$\sum_{i=1}^{n} a_i X_i + b,$$

où a_i, b sont des coefficients non aléatoires. Nous allons démontrer que

$$M\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i M[X_i] + b, \tag{10.2.6}$$

c'est-à-dire que l'espérance mathématique d'une fonction linéaire est égale à cette même fonction linéaire des espérances mathématiques des arguments.

Démonstration. Appliquant le théorème d'addition des espérances mathématiques et mettant en facteur la variable non aléatoire devant le signe de l'espérance mathématique on a :

$$M\left[\sum_{i=1}^{n} a_i X_i + b\right] = M\left[\sum_{i=1}^{n} a_i X_i\right] + M[b] =$$

$$= \sum_{i=1}^{n} M[a_i X_i] + b = \sum_{i=1}^{n} a_i M[X_i] + b.$$

7. *Variance de la somme de variables aléatoires*

La variance de la somme de deux variables aléatoires est égale à la somme de leurs variances à laquelle s'ajoute la double covariance :

$$D[X+Y] = D[X] + D[Y] + 2K_{xy}. \tag{10.2.7}$$

Démonstration. Notons :

$$Z = X + Y. \tag{10.2.8}$$

En vertu du théorème d'addition des espérances mathématiques on a :

$$m_z = m_x + m_y. \tag{10.2.9}$$

Passons des variables X, Y, Z aux variables centrées correspondantes $\dot{X}$, $\dot{Y}$, $\dot{Z}$. En retranchant membre à membre l'égalité (10.2.9) de l'égalité (10.2.8) on a :

$$\dot{Z} = \dot{X} + \dot{Y}.$$

Par définition de la variance

$$D[X+Y] = D[Z] = M[\dot{Z}^2] = M[\dot{X}^2] + 2M[\dot{X}\dot{Y}] + \\ + M[\dot{Y}^2] = D[X] + 2K_{xy} + D[Y],$$

ce qu'il fallait démontrer.

La formule (10.2.7) pour la variance peut être généralisée à un nombre quelconque de termes, soit :

$$D[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} D[X_i] + 2\sum_{i<j} K_{ij}, \qquad (10.2.10)$$

où K_{ij} est la covariance des variables X_i, X_j; la notation $i < j$ sous le signe Σ désigne que la sommation s'étend à toutes les combinaisons possibles deux à deux des variables aléatoires $(X_1, X_2, \ldots, X_n)$.

La démonstration est analogue à celle donnée précédemment et découle de la formule du carré d'un polynôme.

On peut écrire la formule (10.2.10) différemment, soit :

$$D[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij}, \qquad (10.2.11)$$

où la double somme s'étend à tous les éléments de la matrice de variances-covariances du système $(X_1, X_2, \ldots, X_n)$.

Si toutes les variables aléatoires $(X_1, X_2, \ldots, X_n)$ du système ne sont pas corrélées deux à deux (ce qui signifie que $K_{ij} = 0$ pour $i \neq j$), la formule (10.2.10) devient :

$$D[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} D[X_i], \qquad (10.2.12)$$

c'est-à-dire que *la variance de la somme de variables aléatoires non corrélées est égale à la somme des variances des composantes.*

Cette règle est dite *théorème d'addition des variances.*

8. *Variance d'une fonction linéaire*

Considérons une fonction linéaire de plusieurs variables aléatoires

$$\sum_{i=1}^{n} a_i X_i + b,$$

où a_i, b sont des grandeurs non aléatoires.

Nous allons démontrer que la variance de cette fonction linéaire est donnée par la formule:

$$D\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 D[X_i] + 2\sum_{i<j} a_i a_j K_{ij}, \qquad (10.2.13)$$

où K_{ij} est la covariance des variables aléatoires X_i, X_j.

Démonstration. Notons:

$$Y_i = a_i X_i,$$

on a alors:

$$\sum_{i=1}^{n} a_i X_i + b = \sum_{i=1}^{n} Y_i + b. \qquad (10.2.14)$$

En appliquant au second membre de l'expression (10.2.14) la formule (10.2.10) et vu que $D[b] = 0$, on a:

$$D\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} D[Y_i] + 2\sum_{i<j} K_{ij}^{(y)} =$$

$$= \sum_{i=1}^{n} a_i^2 D[X_i] + 2\sum_{i<j} K_{ij}^{(y)}, \qquad (10.2.15)$$

$K_{ij}^{(y)}$ est ici la covariance des variables Y_i, Y_j donnée par la formule:

$$K_{ij}^{(y)} = M[\overset{\circ}{Y}_i \overset{\circ}{Y}_j].$$

Calculons cette quantité:

$$\overset{\circ}{Y}_i = Y_i - m_{y_i} = a_i X_i - a_i m_{x_i} = a_i \overset{\circ}{X}_i;$$

d'une manière analogue:

$$\overset{\circ}{Y}_j = a_j \overset{\circ}{X}_j,$$

d'où

$$K_{ij}^{(y)} = M[a_i a_j \overset{\circ}{X}_i \overset{\circ}{X}_j] = a_i a_j M[\overset{\circ}{X}_i \overset{\circ}{X}_j] = a_i a_j K_{ij}.$$

En substituant cette expression dans (10.2.15) on aboutit à la formule (10.2.13).

Dans le cas particulier où toutes les variables $(X_1, X_2, \ldots, X_n)$ ne sont pas corrélées la formule (10.2.13) devient:

$$D\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 D[X_i], \qquad (10.2.16)$$

c'est-à-dire que *la variance d'une fonction linéaire de variables aléatoires non corrélées est égale à la somme des produits des carrés des coefficients par les variances des arguments correspondants* *).

*) Des variables indépendantes étant toujours non corrélées, toutes les propriétés démontrées dans ce paragraphe pour les variables non corrélées seront également vraies pour les variables indépendantes.

9. *Espérance mathématique du produit de variables aléatoires*

L'espérance mathématique du produit de deux variables aléatoires est égale au produit de leurs espérances mathématiques plus la covariance:

$$M[XY] = M[X]\,M[Y] + K_{xy}. \qquad (10.2.17)$$

Démonstration. Par définition de la covariance on a:

$$K_{xy} = M[\overset{\circ}{X}\overset{\circ}{Y}] = M[(X - m_x)(Y - m_y)],$$

où

$$m_x = M[X]; \quad m_y = M[Y].$$

Transformons cette équation en faisant usage des propriétés de l'espérance mathématique:

$$K_{xy} = M[(X - m_x)(Y - m_y)] =$$
$$= M[XY] - m_x M[Y] - m_y M[X] - m_x m_y = M[XY] - M[X]\,M[Y],$$

ce qui de toute évidence est équivalent à la formule (10.2.17).

Si les variables aléatoires (X, Y) ne sont pas corrélées ($K_{xy} = 0$), la formule (10.2.17) devient:

$$M[XY] = M[X]\,M[Y], \qquad (10.2.18)$$

c'est-à-dire que *l'espérance mathématique du produit de deux variables aléatoires non corrélées est égale au produit de leurs espérances mathématiques.*

Cette formule est connue sous le nom de *théorème de multiplication des espérances mathématiques.*

La formule (10.2.17) n'est rien d'autre que l'expression du second moment mixte centré du système exprimé en fonction du second moment mixte initial et des espérances mathématiques:

$$K_{xy} = \mu_{11} = \alpha_{11} - m_x m_y. \qquad (10.2.19)$$

Cette expression est souvent utilisée pour le calcul du moment de corrélation, tout comme dans le cas d'une seule variable aléatoire la variance se calcule par le second moment initial et l'espérance mathématique.

Le théorème de multiplication des espérances mathématiques peut être généralisé facilement au cas d'un nombre quelconque de facteurs, seulement pour l'appliquer il ne suffit plus que les variables soient non corrélées, il faut encore que certains moments mixtes d'ordre élevé s'annulent, le nombre de ces moments dépendant du nombre de termes figurant dans le produit. Ces conditions sont tou-

jours remplies lorsque les variables sont indépendantes. Dans ce cas

$$M\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} M\left[X_i\right], \qquad (10.2.20)$$

c'est-à-dire que *l'espérance mathématique du produit de variables aléatoires indépendantes est égale au produit de leurs espérances mathématiques.*

Cette formule se démontre facilement par récurrence.

10. Variance d'un produit de variables aléatoires indépendantes

Nous allons démontrer la formule suivante pour les variables indépendantes X et Y:

$$D[XY] = D[X]\,D[Y] + m_x^2 D[Y] + m_y^2 D[X]. \qquad (10.2.21)$$

D é m o n s t r a t i o n. Notons $XY = Z$; par définition de la variance on a:

$$D[XY[= D[Z] = M[\mathring{Z}^2] = M[(Z - m_z)^2].$$

X et Y étant indépendantes, on a $m_z = m_x m_y$ et

$$D[XY] = M[(XY - m_x m_y)^2] = M[X^2Y^2] - 2m_x m_y M[XY] + m_x^2 m_y^2.$$

Il découle de l'indépendance de X et Y que X^2, Y^2 le sont également *); par conséquent:

$$M[X^2Y^2] = M[X^2]\,M[Y^2], \quad M[XY] = m_x m_y$$

et

$$D[XY] = M[X^2]\,M[Y^2] - m_x^2 m_y^2. \qquad (10.2.22)$$

Mais $M[X^2]$ n'est rien d'autre que le second moment initial de X et par conséquent peut s'exprimer en fonction de la variance comme suit:

$$M[X^2] = D[X] + m_x^2;$$

d'une manière analogue on a:

$$M[Y^2] = D[Y] + m_y^2.$$

En substituant ces expressions dans la formule (10.2.22) et en réduisant les termes semblables on obtient la formule (10.2.21).

---

*) On peut montrer que toutes fonctions de variables aléatoires indépendantes sont également indépendantes.

Dans le cas de multiplication de variables centrées (donc, ayant les espérances mathématiques nulles) la formule (10.2.21) devient :

$$D[\mathring{X}\mathring{Y}] = D[\mathring{X}]D[\mathring{Y}], \qquad (10.2.23)$$

c'est-à-dire que *la variance du produit de variables aléatoires indépendantes centrées est égale au produit de leurs variances.*

11. Moments d'ordre supérieur d'une somme de variables aléatoires

Il y a parfois lieu de calculer les moments d'ordre supérieur d'une somme de variables aléatoires indépendantes. Nous allons démontrer certaines relations utiles.

1) Si les variables aléatoires X, Y sont indépendantes on a :

$$\mu_3[X+Y] = \mu_3[X] + \mu_3[Y]. \qquad (10.2.24)$$

Démonstration.

$$\mu_3[X+Y] = M[(X+Y-m_x-m_y)^3] = M[\{(X-m_x)+(Y-m_y)\}^3] =$$
$$= M[(X-m_x)^3] + 3M[(X-m_x)^2(Y-m_y)] +$$
$$+ 3M[(X-m_x)(Y-m_y)^2] + M[(Y-m_y)^3],$$

d'où en vertu du théorème de multiplication des espérances mathématiques

$$\mu_3[X+Y] = \mu_3[X] + 3\mu_2[X]\mu_1[Y] + 3\mu_1[Y]\mu_2[X] + \mu_3[Y].$$

Le moment centré un μ_1 pour une variable quelconque étant nul, les deux termes centraux de cette expression s'annulent, donc, la formule (10.2.24) se trouve démontrée.

En procédant par récurrence on peut facilement généraliser la relation (10.2.24) au cas d'un nombre quelconque de variables aléatoires indépendantes :

$$\mu_3\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mu_3[X_i]. \qquad (10.2.25)$$

2) Le moment centré quatre d'une somme de deux variables aléatoires indépendantes est donné par la formule :

$$\mu_4[X+Y] = \mu_4[X] + \mu_4[Y] + 6D_xD_y, \qquad (10.2.26)$$

où D_x, D_y sont les variances de X et Y.

La démonstration est analogue à celle du cas précédent.

On peut facilement généraliser par récurrence les formules (10.2.26) au cas d'un nombre quelconque de termes indépendants:

$$\mu_4 [\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mu_4 [X_i] + 6 \sum_{i<j} D_{x_i} D_{x_j}. \qquad (10.2.27)$$

Des relations analogues peuvent, s'il y a lieu, être démontrées pour les moments d'ordres plus élevés.

12. Addition des vecteurs aléatoires non corrélés

Soient dans le plan xOy deux vecteurs aléatoires non corrélés: $\vec{V}_1$ de composantes (X_1, Y_1) et $\vec{V}_2$ de composantes (X_2, Y_2) (fig. 10.2.1).

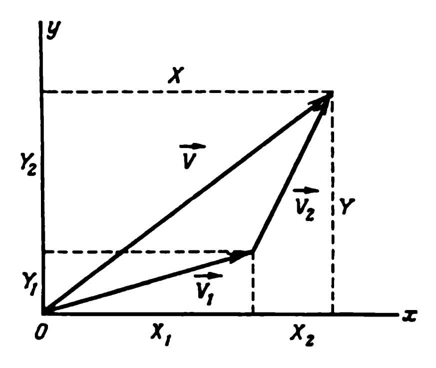


Fig. 10.2.1

Considérons leur somme vectorielle:

$$\vec{V} = \vec{V}_1 + \vec{V}_2,$$

c'est-à-dire le vecteur de composantes:

$$X = X_1 + X_2,$$
$$Y = Y_1 + Y_2.$$

Il y a lieu de trouver les caractéristiques numériques suivantes du vecteur aléatoire $\vec{V}$: les espérances mathématiques m_x, m_y, les variances D_x, D_y et la covariance K_{xy}.

En vertu du théorème d'addition des espérances mathématiques on a:

$$m_x = m_{x_1} + m_{x_2};$$
$$m_y = m_{y_1} + m_{y_2}$$

et en vertu du théorème d'addition des variances:

$$D_x = D_{x_1} + D_{x_2};$$
$$D_y = D_{y_1} + D_{y_2}.$$

Nous allons démontrer maintenant que les covariances s'ajoutent également, c'est-à-dire que

$$K_{xy} = K_{x_1y_1} + K_{x_2y_2}, \qquad (10.2.28)$$

où $K_{x_1y_1}$, $K_{x_2y_2}$ sont les covariances des composantes de chacun des vecteurs $\vec{V}_1$ et $\vec{V}_2$.

D é m o n s t r a t i o n. Par définition de la covariance on a

$$K_{xy} = M[\dot{X}\dot{Y}] = M[(\dot{X}_1 + \dot{X}_2)(\dot{Y}_1 + \dot{Y}_2)] =$$
$$= M[\dot{X}_1\dot{Y}_1] + M[\dot{X}_2\dot{Y}_1] + M[\dot{X}_1\dot{Y}_2] + M[\dot{X}_2\dot{Y}_2]. \qquad (10.2.29)$$

Les vecteurs $\vec{V}_1$, $\vec{V}_2$ étant non corrélés, les deux termes du milieu dans la formule (10.2.29) sont nuls; les deux termes restants sont $K_{x_1y_1}$ et $K_{x_2y_2}$, donc la formule (10.2.28) se trouve démontrée.

La formule (10.2.28) est parfois appelée « théorème d'addition des covariances ».

Ce théorème peut facilement être généralisé au cas d'un nombre quelconque de termes. Soient deux systèmes de variables aléatoires non corrélées, c'est-à-dire deux vecteurs aléatoires non corrélés à n dimensions:

$\vec{X}$ de composantes $X_1, X_2, \ldots, X_n$,
$\vec{Y}$ de composantes $Y_1, Y_2, \ldots, Y_n$,

la matrice de corrélation de la somme vectorielle:

$$\vec{Z} = \vec{X} + \vec{Y}$$

a pour éléments les sommes des éléments des matrices de corrélation des composantes:

$$K_{ij}^{(z)} = K_{ij}^{(x)} + K_{ij}^{(y)}, \qquad (10.2.30)$$

où $K_{ij}^{(z)}$, $K_{ij}^{(x)}$, $K_{ij}^{(y)}$ sont respectivement les covariances des variables (Z_i, Z_j); (X_i, X_j); (Y_i, Y_j).

La formule (10.2.30) est vraie tant pour $i = j$ que pour $i \neq j$.

En effet, les composantes du vecteur $\vec{Z}$ sont:

$$Z_1 = X_1 + Y_1;$$
$$Z_2 = X_2 + Y_2;$$
$$\cdots\cdots$$
$$Z_n = X_n + Y_n.$$

En vertu du théorème d'addition des variances:

$$D_{z_i} = D_{x_i} + D_{y_i},$$

ou, en utilisant des notations différentes:

$$K_{ii}^{(z)} = K_{ii}^{(x)}K + K_{ii}^{(y)} .$$

En vertu de théorème d'addition des covariances pour $i \neq j$ on a:

$$K_{ij}^{(z)} = K_{ij}^{(x)} + K_{ij}^{(y)}.$$

En mathématiques on appelle *somme de deux matrices* la matrice dont les éléments sont obtenus par addition des éléments correspondants de ces matrices. En utilisant cette terminologie on peut dire que la matrice de corrélation de la somme de deux vecteurs aléatoires non corrélés est égale à la somme des matrices de corrélation des composantes:

$$\| K_{ij}^{(z)} \| = \| K_{ij}^{(x)} \| + \| K_{ij}^{(y)} \|. \qquad (10.2.31)$$

Par analogie avec les cas précédents on appelle cette règle « théorème d'addition des matrices de corrélation ».

10.3. Applications des théorèmes sur les caractéristiques numériques

Dans ce paragraphe nous allons montrer comment on peut utiliser l'appareil des caractéristiques numériques à la résolution d'un certain nombre de problèmes. Certains de ces problèmes présentent un intérêt théorique par eux-mêmes et seront utilisés dans la suite; d'autres problèmes ne sont donnés qu'à titre d'exemple.

Problème 1. Coefficient de corrélation de variables aléatoires à dépendance linéaire.

Montrer que si les variables aléatoires X et Y sont liées par une relation fonctionnelle linéaire

$$Y = aX + b,$$

leur coefficient de corrélation est égal à $+1$ ou -1, suivant le signe du coefficient a.

Solution. On a:

$$K_{xy} = M[\mathring{X}\mathring{Y}] = M[(X - m_x)(Y - m_y)] =$$
$$= M[(X - m_x)(aX + b - am_x - b)] = aM[(X - m_x)^2] = aD_x,$$

où D_x est la variance de X.

Ecrivons l'expression du coefficient de corrélation:

$$r_{xy}=\frac{K_{xy}}{\sigma_x\sigma_y}. \qquad (10.3.1)$$

Pour calculer σ_y on recherche la variance de Y:

$$D_y=D[aX+b]=a^2D_x,$$
$$\sigma_y=|a|\sigma_x.$$

La substitution dans la formule (10.3.1) donne:

$$r_{xy}=\frac{aD_x}{|a|\sigma_x^2}=\frac{a}{|a|}.$$

La grandeur $\frac{a}{|a|}$ est égale à $+1$ lorsque a est positif et à -1 lorsque a est négatif, c'est ce qu'il fallait démontrer.

Problème 2. Limites du coefficient de corrélation.

Montrer que pour des variables aléatoires quelconques on a:

$$|r_{xy}|\leqslant 1.$$

Solution. Soit la variable aléatoire

$$Z=\sigma_yX\pm\sigma_xY,$$

où σ_x, σ_y sont les écarts quadratiques moyens de X et Y. Nous allons calculer la variance de Z. Selon la formule (10.2.13) on a:

$$D_z=\sigma_y^2D_x+\sigma_x^2D_y\pm 2\sigma_x\sigma_yK_{xy}$$

ou:

$$D_z=2\sigma_x^2\sigma_y^2\pm 2\sigma_x\sigma_yK_{xy}.$$

La variance d'une variable aléatoire quelconque ne pouvant être négative, on a:

$$2\sigma_x^2\sigma_y^2\pm 2\sigma_x\sigma_yK_{xy}\geqslant 0,$$

ou

$$\sigma_x\sigma_y\pm K_{xy}\geqslant 0,$$

d'où:

$$|K_{xy}|\leqslant\sigma_x\sigma_y,$$

par conséquent:

$$|r_{xy}|\leqslant 1,$$

ce qu'il fallait démontrer.

Problème 3. Projection d'un point aléatoire d'un plan sur une droite quelconque.

Soit un point aléatoire de coordonnées (X, Y) dans un plan (fig. 10.3.1). Projetons ce point sur l'axe Oz menée par l'origine des coordonnées et faisant avec l'axe Ox un angle α. La projection du

point (X, Y) sur l'axe Oz est également un point aléatoire ; sa distance Z à l'origine des coordonnées est une variable aléatoire. Trouver l'espérance mathématique et la variance de Z.

Solution. On a:

$$Z = X \cos \alpha + Y \sin \alpha.$$

Z étant une fonction linéaire des arguments X et Y,

$$m_z = m_x \cos \alpha + m_y \sin \alpha;$$

$$D_z = D_x \cos^2 \alpha + D_y \sin^2 \alpha + 2K_{xy} \sin \alpha \cos \alpha =$$
$$= D_x \cos^2 \alpha + D_y \sin^2 \alpha + K_{xy} \sin 2\alpha,$$

D_x, D_y, K_{xy} étant respectivement les variances et la covariance des variables (X, Y).

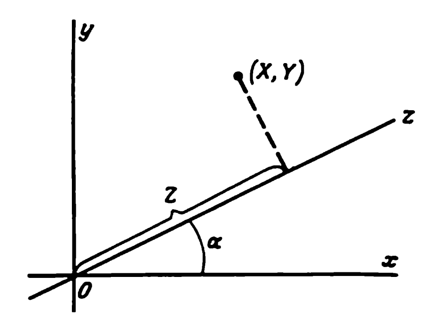


Fig. 10.3.1

En passant aux écarts quadratiques moyens on a:

$$\sigma_z^2 = \sigma_x^2 \cos^2 \alpha + \sigma_y^2 \sin^2 \alpha + r_{xy}\sigma_x\sigma_y \sin 2\alpha. \quad (10.3.2)$$

Dans le cas de variables aléatoires non corrélées (pour $r_{xy} = 0$) on obtient:

$$\sigma_z^2 = \sigma_x^2 \cos^2 \alpha + \sigma_y^2 \sin^2 \alpha. \quad (10.3.3)$$

Problème 4. Espérance mathématique du nombre de réalisation d'un événement dans plusieurs expériences.

Supposons qu'on effectue n expériences, dans chacune desquelles l'événement A peut apparaître ou ne pas apparaître. La probabilité de réalisation de l'événement A dans l'i-ème expérience est égale à p_i. Trouver l'espérance mathématique du nombre de réalisations de l'événement.

Solution. Soit X la variable aléatoire discrète représentant le nombre de réalisations de l'événement dans toute la série d'expé-

riences. Il est évident que:

$$X = X_1 + X_2 + \ldots + X_n,$$

ou plus brièvement:

$$X = \sum_{i=1}^{n} X_i,$$

où X_i est le nombre de réalisations de l'événement A dans l'i-ème expérience; $i = 1, \ldots, n$.

Chacun des nombres X_i est une variable aléatoire discrète pouvant prendre l'une des deux valeurs 0 et 1. Le tableau de répartition de X_i est:

$$\frac{0 \mid 1}{q_i \mid p_i}, \tag{10.3.4}$$

où $q_i = 1 - p_i$ est la probabilité de non-réalisation de l'événement A dans l'i-ème expérience.

En vertu du théorème d'addition des espérances mathématiques:

$$m_x = M[X] = \sum_{i=1}^{n} m_{x_i}, \tag{10.3.5}$$

où m_{x_i} est l'espérance mathématique de X_i.

Calculons m_{x_i}. Par définition de l'espérance mathématique on a:

$$m_{x_i} = 0 \cdot q_i + 1 \cdot p_i = p_i.$$

En substituant ce résultat dans la formule (10.3.5) on obtient:

$$m_x = \sum_{i=1}^{n} p_i, \tag{10.3.6}$$

c'est-à-dire que *l'espérance mathématique du nombre de réalisations d'un événement dans une série d'expériences est égale à la somme des probabilités de l'événement dans chaque expérience particulière.*

En particulier, lorsque les conditions des expériences sont les mêmes et

$$p_1 = p_2 = \ldots = p_n = p,$$

la formule (10.3.5) devient:

$$m_x = np. \tag{10.3.7}$$

Comme le théorème d'addition des esperances mathématiques est applicable à des variables aléatoires quelconques, tant dépendantes qu'indépendantes, les formules (10.3.6) et (10.3.7) sont indifféremment vraies pour des expériences liées et indépendantes.

Le théorème démontré est souvent appliqué dans la théorie du tir, lorsqu'il y a lieu de trouver le nombre moyen d'atteintes pour plu-

sieurs coups, tant dépendants qu'indépendants. *L'espérance mathématique du nombre d'atteintes pour plusieurs coups est égale à la somme des probabilités d'atteinte pour différents coups.*

Problème 5. Variance du nombre de réalisations d'un événement dans plusieurs expériences indépendantes.

On effectue n expériences indépendantes chacune desquelles peut réaliser ou ne pas réaliser l'événement A ; la probabilité de réalisation de l'événement A dans l'i-ème expérience est égale à p_i. Trouver la variance et l'écart quadratique moyen du nombre de réalisations de l'événement A.

Solution. Considérons la variable aléatoire X représentant le nombre de réalisations de l'événement A, que nous écrirons comme dans le problème précédent comme la somme

$$X = \sum_{i=1}^{n} X_i,$$

où X_i est le nombre de réalisations de l'événement A dans l'i-ème expérience.

Les expériences étant indépendantes, les variables aléatoires X_1, X_2, . . ., X_n le sont également et on peut leur appliquer le théorème d'addition des variances:

$$D_x = \sum_{i=1}^{n} D_{x_i}.$$

Calculer maintenant la variance de X_i. Le tableau de répartition (10.3.4) donne :

$$D_{x_i} = (0 - p_i)^2 q_i + (1 - p_i)^2 p_i = p_i q_i,$$

d'où

$$D_x = \sum_{i=1}^{n} p_i q_i, \tag{10.3.8}$$

c'est-à-dire que *la variance du nombre d'apparitions d'un événement dans plusieurs expériences indépendantes est égale à la somme des produits des probabilités de réalisation et de non-réalisation de l'événement dans chacune des expériences.*

La formule (10.3.8) donne l'écart quadratique moyen du nombre de réalisations de l'événement A :

$$\sigma_x = \sqrt{\sum_{i=1}^{n} p_i q_i}. \tag{10.3.9}$$

Pour des conditions d'expériences constantes, lorsque $p_1 = p_2 = \ldots = p_n = p$, les formules (10.3.8) et (10.3.9) se

simplifient et on a:

$$\left.\begin{aligned} D_x &= npq, \\ \sigma_x &= \sqrt{npq}. \end{aligned}\right\} \qquad (10.3.10)$$

Problème 6. Variance du nombre d'apparitions d'un événement dans une série d'expériences dépendantes.

Soient n expériences dépendantes pouvant réaliser ou ne pas réaliser chacune l'événement A ; la probabilité de cet événement dans l'i-ème expérience est égale à p_i ($i = 1, 2, \ldots, n$). Trouver la variance du nombre de réalisations de cet événement.

Solution. Pour résoudre le problème posé nous allons de nouveau écrire le nombre de réalisations de l'événement X comme la somme:

$$X = \sum_{i=1}^{n} X_i, \qquad (10.3.11)$$

où

$$X_i = \begin{cases} 1, & \text{si l'événement } A \text{ a été réalisé par l'}i\text{-ème expérience,} \\ 0, & \text{si l'événement } A \text{ n'a pas été réalisé par l'}i\text{-ème expérience.} \end{cases}$$

Les expériences n'étant pas indépendantes, il ne suffit plus de connaître les probabilités

$$p_1, p_2, \ldots, p_n$$

de réalisation de l'événement A par la première, la seconde, la troisième, etc. expérience. Il faut de plus se donner la manière dont les expériences dépendent les unes des autres. Il se trouve que pour résoudre ce problème il suffit de connaître la probabilité P_{ij} de réalisation conjointe de l'événement A tant dans l'i-ème que la j-ième expérience; $P((X_i = 1)(X_j = 1)) = P_{ij}$. Supposons que cette probabilité soit donnée. Nous allons appliquer à l'expression (10.3.11) le théorème d'addition des variances [formule (10.2.10)]:

$$D_x = \sum_{i=1}^{n} D_{x_i} + 2 \sum_{i<j} K_{ij}, \qquad (10.3.12)$$

où K_{ij} est la covariance de X_i et X_j:

$$K_{ij} = M[\mathring{X}_i \mathring{X}_j].$$

La formule (10.2.19) donne:

$$K_{ij} = M[X_i X_j] - m_{x_i} m_{x_j} = M[X_i X_j] - p_i p_j. \qquad (10.3.13)$$

Considérons la variable aléatoire $X_i X_j$. Il est évident qu'elle est égale à zéro si l'une des variables X_i, X_j est égale à zéro, c'est-à-dire si au moins l'une des expériences (l'i-ème ou la j-ième) n'a

pas réalisé l'événement A. Pour que la quantité X_iX_j soit égale à l'unité il faut que l'événement A apparaisse dans les deux expériences. Cette probabilité est égale à P_{ij}. Par conséquent

$$M[X_iX_j] = P_{ij}$$

et

$$K_{ij} = P_{ij} - p_ip_j.$$

En substituant cette expression dans la formule (10.3.12) on obtient :

$$D_x = \sum_{i=1}^{n} p_iq_i + 2\sum_{i<j}(P_{ij} - p_ip_j). \qquad (10.3.14)$$

La formule (10.3.14) donne la variance du nombre de réalisations de l'événement dans des expériences dépendantes. Le premier terme du second membre est la variance du nombre de réalisations de l'événement dans les expériences indépendantes, et le second, la correction due à la dépendance des expériences. Si la probabilité P_{ij} est égale à p_ip_j, cette correction est égale à zéro. Si la probabilité P_{ij} est supérieure à p_ip_j, ceci signifie que la probabilité conditionnelle de réalisation de l'événement A par la j-ième expérience à condition que l'i-ème expérience l'a réalisé, est supérieure à la probabilité simple p_j (inconditionnelle) de réalisation de l'événement dans la j-ème expérience (entre les réalisations de l'événement dans l'i-ième et la j-ième expérience il y a une corrélation positive). S'il en est ainsi pour tout couple d'expériences, le terme correcteur dans la formule (10.3.14) est positif et la variance du nombre de réalisations de l'événement dans des expériences dépendantes est supérieure à la variance dans le cas des expériences indépendantes.

Si la probabilité P_{ij} est inférieure à p_ip_j (entre les réalisations de l'événement dans l'i-ème et la j-ième expérience il y a une corrélation négative), le terme correspondant est négatif. S'il en est ainsi pour tout couple d'expériences, la variance du nombre de réalisations de l'événement dans des expériences dépendantes est inférieure à la variance dans le cas des expériences indépendantes.

Considérons le cas particulier où $p_1 = p_2 = \ldots = p_n = p$; $P_{11} = P_{12} = \ldots = P$, c'est-à-dire que les conditions de toutes les expériences sont les mêmes. La formule (10.3.14) devient alors :

$$D_x = npq + 2\sum_{i<j}(P - p^2) = npq + n(n-1)(P - p^2), \qquad (10.3.15)$$

où P est la probabilité conjointe de réalisation de l'événement A dans deux quelconques expériences.

Deux cas particuliers sont intéressants :

1. La réalisation de l'événement A par n'importe laquelle des expériences entraîne obligatoirement sa réalisation par toutes les

autres expériences. On a alors $P = p$ et la formule (10.3.15) devient :

$$D_x = npq + n(n-1)(p-p^2) = npq + n(n-1)pq = n^2pq.$$

2. La réalisation de l'événement A par une expérience quelconque rend impossible sa réalisation par toutes les autres expériences. On a alors $P = 0$ et la formule (10.3.15) devient :

$$D_x = npq - n(n-1)p^2 = np[q-(n-1)p] = np(1-np).$$

Problème 7. Espérance mathématique du nombre d'unités mises à un état donné.

Dans les applications pratiques on a souvent à résoudre le problème suivant. Un certain groupe est formé de n unités sur lesquelles on exerce une certaine action. Par suite de cette action chacune des unités peut se trouver dans un certain état S. La probabilité pour l'i-ème unité de se trouver dans l'état S est égale à p_i. Trouver l'espérance mathématique du nombre d'unités qui, par suite de l'action sur tout le groupe, se trouveront dans l'état S.

Solution. Faisons correspondre à chacune des unités la variable aléatoire X_i pouvant prendre les valeurs 0 et 1 :

$$X_i = \begin{cases} 1 \text{ si l'}i\text{-ème unité se trouve dans l'état } S; \\ 0 \text{ si l'}i\text{-ème unité n'est pas dans l'état } S. \end{cases}$$

Soit X le nombre d'unités mises à l'état S, il est évident que

$$X = \sum_{i=1}^{n} X_i.$$

En vertu du théorème d'addition des espérances mathématiques on a :

$$m_x = \sum_{i=1}^{n} m_{x_i}.$$

L'espérance mathématique de chacune des variables X_i est connue et égale à :

$$m_{x_i} = p_i.$$

Par conséquent :

$$m_x = \sum_{i=1}^{n} p_i, \qquad (10.3.16)$$

c'est-à-dire que l'espérance mathématique du nombre d'unités mises à l'état S est égale à la somme des probabilités de réduction à cet état de chacune des unités.

Notons que pour être vraie la formule démontrée n'exige obligatoirement pas que les unités passent à l'état S indépendamment les unes des autres. De plus, la formule est vraie pour n'importe quel genre d'action.

Problème 8. Variance du nombre d'unités mises à un état donné.

Si pour les conditions du problème précédent la réduction de chacune des unités à l'état S se produit indépendamment de toutes les autres, le théorème d'addition des variances, appliqué à la quantité

$$X = \sum_{i=1}^{n} X_i,$$

donne la variance du nombre d'unités mises à l'état S:

$$D_x = \sum_{i=1}^{n} D_{x_i} = \sum_{i=1}^{n} p_i q_i, \qquad q_i = 1 - p_i. \tag{10.3.17}$$

Par contre, si l'action sur les unités est telle que le passage à l'état S pour certaines unités se réalise conjointement, la variance du nombre d'unités mises à l'état S s'exprime par la formule suivante (cf. formule 10.3.14):

$$D_x = \sum_{i=1}^{n} p_i q_i + 2 \sum_{i<j} (P_{ij} - p_i p_j), \tag{10.3.18}$$

où P_{ij} est la probabilité pour les unités i et j de passer conjointement à l'état S à la suite d'une certaine action.

Problème 9. Espérance mathématique du nombre d'expériences avant la k-ième apparition de l'événement.

Soient une série d'expériences indépendantes dans chacune desquelles l'événement A peut apparaître ou ne pas apparaître avec une probabilité p. Les expériences durent tant que l'événement A n'apparaisse k fois, puis les expériences sont arrêtées. Trouver l'espérance mathématique, la variance et l'écart quadratique moyen du nombre d'expériences X qui seront éventuellement effectuées.

Solution. Dans l'exemple 3 du § 5.7 on a trouvé l'espérance mathématique et la variance du nombre d'expériences précédant la première réalisation de l'événement A:

$$m = \frac{1}{p}; \qquad D = \frac{q}{p^2},$$

où p est la probabilité de réalisation de l'événement dans l'une des expériences,

$q = 1 - p$ la probabilité de sa non-réalisation.

Considérons la variable aléatoire X représentant le nombre d'expériences précédant la réalisation de l'événement A. Elle peut s'écrire comme la somme:

$$X = X_1 + X_2 + \ldots + X_k,$$

où X_1 est le nombre d'expériences précédant la première réalisation de l'événement A,

X_2, le nombre d'expériences séparant la première et la seconde réalisation de l'événement A (seconde réalisation incluse),

. . . .

X_k, le nombre d'expériences séparant la $(k-1)$-ième et la k-ième réalisation de l'événement A (k-ième réalisation incluse).

Il est évident que les variables $X_1, X_2, \ldots, X_k$ sont indépendantes et sont réparties toutes suivant la même loi que X_1 (nombre d'expériences précédant la première réalisation de l'événement), les caractéristiques numériques de cette loi étant

$$m_{x_i} = \frac{1}{p}; \qquad D_{x_i} = \frac{q}{p^2}.$$

Les théorèmes d'addition des espérances mathématiques et des variances donnent:

$$\left.\begin{aligned} m_x &= \sum_{i=1}^{k} m_{x_i} = \frac{k}{p}; \\ D_x &= \sum_{i=1}^{k} D_{x_i} = \frac{kq}{p^2}; \\ \sigma_x &= \frac{\sqrt{kq}}{p}. \end{aligned}\right\} \qquad (10.3.19)$$

P r o b l è m e 10. D é p e n s e s m o y e n n e s a v a n t d' a t t e i n d r e u n c e r t a i n r é s u l t a t.

Dans le problème précédent nous avons envisagé le cas où une série d'expériences est entreprise dans le but d'atteindre un certain résultat — k réalisations de l'événement A, qui dans chaque expérience a la même probabilité. Ce problème est un cas particulier d'un autre problème où une série d'expériences est effectuée dans le but d'atteindre un résultat quelconque B, dont la probabilité augmente avec le nombre d'expériences n suivant une loi quelconque $P(n)$. Supposons que chaque expérience coûte a. Trouver l'espérance mathématique des dépenses.

S o l u t i o n. Pour résoudre le problème posé on suppose que le nombre d'expériences n'est limité en rien et que celles-ci sont prolongées après avoir obtenu le résultat B. Certaines de ces expériences seront alors superflues. On dit qu'une expérience est nécessaire, si elle a lieu lorsque le résultat B n'est pas atteint, et superflue si elle a lieu lorsque le résultat B est déjà atteint.

On fait correspondre à chaque i-ème expérience la quantité X_i, égale à zéro ou à l'unité suivant que cette expérience s'est trou-

vée être indispensable ou superflue. Posons donc

$$X_i = \begin{cases} 1 \text{ si l'expérience s'est trouvée être indispensable;} \\ 0 \text{ si elle s'est trouvée être superflue.} \end{cases}$$

Considérons la variable aléatoire X représentant le nombre d'expériences à faire pour obtenir le résultat B. Il est évident qu'elle peut s'écrire comme la somme:

$$X = X_1 + X_2 + \ldots + X_i + \ldots \qquad (10.3.20)$$

Parmi les termes de la somme (10.3.20) le premier, X_1, n'est pas aléatoire, il est toujours égal à l'unité (la première expérience est toujours indispensable). Chacun des autres termes peut aléatoirement prendre l'une des valeurs 0 ou 1. Etablissons le tableau de répartition de la variable aléatoire X_i $(i > 1)$, on a:

0	1
$P(i-1)$	$1-P(i-1)$

(10.3.21)

où $P(i-1)$ est la probabilité d'obtenir le résultat B après $i-1$ expériences.

En effet, si le résultat B a déjà été obtenu lors des $i-1$ expériences précédentes, on a $X_i = 0$ (l'i-ème expérience est superflue); si le résultat n'a pas été obtenu, on a $X_i = 1$ (l' i-ème expérience est indispensable).

Nous allons trouver maintenant l'espérance mathématique de X_i. Le tableau de répartition (10.3.21) donne:

$$m_{x_i} = 0 \cdot P(i-1) + 1 \cdot [1 - P(i-1)] = 1 - P(i-1).$$

Il est facile de voir que la même formule sera également vraie pour $i = 1$ car $P(0) = 0$.

Nous allons appliquer à l'expression (10.3.20) le théorème d'addition des espérances mathématiques. On a:

$$m_x = \sum_{i=1}^{\infty} m_{x_i} = \sum_{i=1}^{\infty} [1 - P(i-1)],$$

ou, en introduisant la désignation $i - 1 = k$:

$$m_x = \sum_{k=0}^{\infty} [1 - P(k)]. \qquad (10.3.22)$$

En prenant le produit de la valeur obtenue de m_x par le coût a de chaque expérience on trouve la somme moyenne à dépenser pour arriver au résultat B:

$$S_B = a \sum_{k=0}^{\infty} [1 - P(k)]. \qquad (10.3.23)$$

Cette formule a été trouvée en supposant que le coût de chaque expérience est le même. S'il n'en est pas ainsi, on peut utiliser une autre méthode, dans laquelle le coût total S_B est la somme des coûts des différentes expériences pouvant prendre deux valeurs: a_i si l'i-ème expérience est indispensable et zéro si elle est superflue. La dépense moyenne S_B est:

$$S_B = \sum_{k=0}^{\infty} a_k [1 - P(k)]. \tag{10.3.24}$$

Problème 11. Espérance mathématique de la somme d'un nombre aléatoire de composantes aléatoires.

Dans de nombreuses applications de la théorie des probabilités on rencontre des sommes aléatoires, dans lesquelles le nombre de composantes est inconnu d'avance, c'est-à-dire aléatoire.

On se propose de résoudre le problème suivant. Soit une variable aléatoire Z somme de Y variables aléatoires:

$$Z = \sum_{i=1}^{Y} X_i, \tag{10.3.25}$$

Y étant également une variable aléatoire. Supposons que l'on connaisse l'espérance mathématique m_{x_i} de toutes les composantes:

$$m_{x_i} = M[X_i],$$

et que Y ne dépende d'aucune des variables X_i.

Trouver l'espérance mathématique de Z.

Solution. Le nombre de composantes de la somme est une variable aléatoire discrète. Supposons que l'on connaisse le tableau de répartition:

y_k	1	2	...	k	...
p_k	p_1	p_2	...	p_k	...

où p_k est la probabilité pour que Y prenne la valeur k. Soit $Y = k$, dans ce cas l'espérance mathématique de Z (espérance mathématique conditionnelle) est:

$$M[Z \mid k] = \sum_{i=1}^{k} m_{x_i}. \tag{10.3.26}$$

Nous allons appliquer maintenant la formule de l'espérance mathématique marginale, à cet effet on multiplie chaque espérance mathématique conditionnelle par la probabilité de l'hypothèse correspondante p_k, on a alors la somme:

$$M[Z] = \sum_k p_k \sum_{i=1}^{k} m_{x_i}. \tag{10.3.27}$$

Le cas où toutes les variables aléatoires $X_1, X_2, \ldots$ ont même espérance mathématique

$$m_{x_1} = m_{x_2} = \ldots = m_x$$

est particulièrement intéressant.

La formule (10.3.26) devient alors:

$$M[Z \mid k] = \sum_{i=1}^{k} m_x = k m_x$$

et

$$M[Z] = m_x \sum_k k p_k. \qquad (10.3.28)$$

La somme dans l'expression (10.3.28) n'est rien d'autre que l'espérance mathématique de Y:

$$m_y = \sum_k k p_k,$$

alors:

$$m_z = M[Z] = m_x \cdot m_y, \qquad (10.3.29)$$

c'est-à-dire que l'espérance mathématique de la somme d'un nombre aléatoire de composantes aléatoires de mêmes valeurs moyennes (si toutefois le nombre de composantes ne dépend pas de ces valeurs) est égale au produit des valeurs moyennes de chacune des composantes par le nombre moyen de composantes.

Remarquons de nouveau que le résultat obtenu est vrai tant pour le cas des composantes $X_1, X_2, \ldots$ indépendantes que dépendantes, pourvu que le nombre de composantes Y n'en dépende pas.

Nous allons terminer le présent chapitre par l'examen d'un certain nombre d'exemples puisés dans différents domaines, pour illustrer comment utiliser les méthodes des caractéristiques numériques et les théorèmes démontrés plus haut pour les problèmes courants.

Exemple 1. On jette 10 fois une pièce de monnaie. Trouver l'espérance mathématique et l'écart quadratique moyen du nombre X de faces.

Solution. Les formules (10.3.7) et (10.3.10) donnent:

$$m_x = 10 \cdot 0{,}5 = 5; \quad D_x = 10 \cdot 0{,}5 \cdot 0{,}5 = 2{,}5; \quad \sigma_x = 1{,}58.$$

Exemple 2. Cinq coups indépendants sont tirés sur une cible ronde de 20 cm de diamètre. On vise le centre de la cible, il n'y a pas d'erreur systématique, la dispersion est circulaire, l'écart quadratique moyen $\sigma = 16$ cm. Trouver l'espérance mathématique et l'écart quadratique moyen du nombre de coups réussis.

Solution. La probabilité d'atteindre la cible avec un coup peut être calculée par la formule (9.4.5):

$$p = 1 - e^{-\frac{k^2}{2}} = 1 - e^{-0{,}625} \approx 0{,}465.$$

On obtient à partir des formules (10.3.7) et (10.3.10):

$$m_x = 5p \approx 2{,}32; \quad D_x = 5p(1-p) \approx 1{,}25; \quad \sigma_x \approx 1{,}12.$$

Exemple 3. Un garage dessert 40 automobiles et 70 camions. La probabilité pour que, durant un an, une automobile soit en réparation est égale à $p_1 = 0{,}6$; la probabilité pour que, durant un an, un camion soit en réparation est égale à $p_2 = 0{,}3$. Les voitures sont réparées indépendamment les unes des autres. Trouver l'espérance mathématique, la variance et l'écart quadratique moyen des grandeurs suivantes:
du nombre X d'automobiles en réparation durant un an;
du nombre Y de camions en réparation durant un an;
du nombre total $Z = X + Y$ de voitures en réparation durant un an.

Solution. D'après les formules (10.3.7) et (10.3.10) on a

$$m_x = 40\cdot 0{,}6 = 24;\quad D_x = 40\cdot 0{,}6\cdot 0{,}4 = 9{,}6;\quad \sigma_x = \sqrt{9{,}6} \approx 3{,}1;$$
$$m_y = 70\cdot 0{,}3 = 21;\quad D_y = 70\cdot 0{,}3\cdot 0{,}7 = 14{,}7;\quad \sigma_y = \sqrt{14{,}7} \approx 3{,}8.$$

En vertu du théorème d'addition des espérances mathématiques:

$$m_z = m_x + m_y = 24 + 21 = 45,$$

et du théorème d'addition des variances:

$$D_z = D_x + D_y = 9{,}6 + 14{,}7 = 24{,}3;\quad \sigma_z = \sqrt{24{,}3} \approx 4{,}9.$$

Exemple 4. Les variables aléatoires X et Y sont les erreurs élémentaires apparaissant à l'entrée d'un appareil. Leurs espérances mathématiques sont $m_x = -2$ et $m_y = 4$, leurs variances $D_x = 4$ et $D_y = 9$ et le coefficient de corrélation de ces erreurs $r_{xy} = -0{,}5$. L'erreur à la sortie de l'appareil est liée à l'erreur d'entrée par la relation fonctionnelle:

$$Z = 3X^2 - 2XY + Y^2 - 3.$$

Trouver l'espérance mathématique de l'erreur à la sortie de l'appareil.

Solution.

$$m_z = M[Z] = 3M[X^2] - 2M[XY] + M[Y^2] - 3.$$

Faisant usage de la relation existant entre les moments initiaux et centrés et de la formule (10.2.17) on a:

$$M[X^2] = \alpha_2[X] = D_x + m_x^2 = 8;$$
$$M[Y^2] = \alpha_2[Y] = D_y + m_y^2 = 25;$$
$$M[XY] = m_x m_y + K_{xy} = m_x m_y + r_{xy}\sigma_x\sigma_y = -11,$$

d'où $m_z = 3\cdot 8 - 2\cdot(-11) + 25 - 3 = 68$.

Exemple 5. Une urne contient trois boules blanches et six noires. On en tire cinq boules dont X blanches. Trouver l'espérance mathématique et la variance de X.

Solution. Chacun des cinq tirages de boules peut être considéré comme une expérience indépendante. La probabilité p de tirer une boule blanche est la même pour toutes les expériences, elle est égale à $3/9 = 1/3$. D'où $m_x = 5\cdot 1/3 = 5/3$.

Pour trouver la variance de X nous allons appliquer la formule (10.3.15). La probabilité P dans cette formule n'est rien d'autre que la probabilité de tirer deux boules blanches lors de deux tirages:

$$P = 3/9\cdot 2/8 = 1/12.$$

La formule (10.3.15) donne:

$$D_x = 5\cdot 1/3\cdot 2/3 + 5\cdot 4\cdot(1/12 - (1/3)^2) = 5/9.$$

Exemple 6. Un radiotélescope durant un temps t scrute l'espace où se trouve un groupe de quatre objectifs. Durant ce laps de temps chacun des

objectifs peut être détecté. La probabilité de détection du premier objectif est $p_1 = 0{,}2$, du second $p_2 = 0{,}25$, du troisième $p_3 = 0{,}35$, du quatrième $p_4 = 0{,}42$. Trouver l'espérance mathématique et l'écart quadratique moyen du nombre d'objectifs qui seront détectés durant le temps t.

S o l u t i o n. Soit X le nombre d'objectifs détectés. En appliquant la formule (10.3.6) on obtient:

$$m_x = \sum_{i=1}^{4} p_i = 0{,}2 + 0{,}25 + 0{,}35 + 0{,}42 = 1{,}22.$$

La formule (10.3.8) donne:

$$D_x = \sum_{i=1}^{4} p_i q_i = 0{,}2\cdot 0{,}8 + 0{,}25\cdot 0{,}75 + 0{,}35\cdot 0{,}65 + 0{,}42\cdot 0{,}58 = 0{,}818;$$

$$\sigma_x = \sqrt{0{,}818} \approx 0{,}904.$$

E x e m p l e 7. On envisage de réaliser un certain nombre de mesures, donnant chacune en cas de succès un bénéfice net égal à X, réparti suivant une loi normale de valeur moyenne $m = 2$ (en unités conventionnelles); le nombre de mesures durant la période considérée est aléatoire et réparti suivant la loi:

y_i	1	2	3	4
p_i	0,2	0,3	0,4	0,1

de plus, ce nombre ne dépend pas du bénéfice offert par les mesures. Trouver le bénéfice moyen éventuel pour toute la période en question.

S o l u t i o n. En se basant sur le problème 11 du présent paragraphe on trouve l'espérance mathématique de bénéfice total Z:

$$m_z = m_x \cdot m_y,$$

où m_x est le bénéfice moyen d'une mesure, m_y le nombre moyen de mesures éventuelles.

On a:

$$m_y = 1\cdot 0{,}2 + 2\cdot 0{,}3 + 3\cdot 0{,}4 + 4\cdot 0{,}1 = 2{,}4;$$
$$m_x = 2;$$
$$m_z = 2\cdot 2{,}4 = 4{,}8.$$

E x e m p l e 8. L'erreur d'un appareil est donnée par la fonction:

$$U = 3Z + 2X - Y - 4, \tag{10.3.30}$$

où X, Y, Z sont les erreurs élémentaires, formant un système de variables aléatoires (vecteur aléatoire).

Le vecteur aléatoire (X, Y, Z) est caractérisé par les espérances mathématiques:

$$m_x = -4; \quad m_y = 1; \quad m_z = 1$$

et la matrice des variances-covariances:

$$\begin{Vmatrix} D_x & K_{xy} & K_{xz} \\ & D_y & K_{yz} \\ & & D_z \end{Vmatrix} = \begin{Vmatrix} 2 & 1 & -1 \\ & 3 & 1 \\ & & 4 \end{Vmatrix}.$$

Trouver l'espérance mathématique, la variance et l'écart quadratique moyen de l'erreur de l'appareil.

S o l u t i o n. La fonction (10.3.30) étant linéaire, on a en vertu des formules (10.2.6) et (10.2.13):

$$m_u = 3m_z + 2m_x - m_y - 4 = -10;$$
$$D_u = 3^2 \cdot D_z + 2^2 \cdot D_x + 1^2 \cdot D_y + 2\,[3 \cdot 2 \cdot K_{xz} + 3\,(-1)\,K_{yz} + 2\,(-1)\,K_{xy}] = 25;$$
$$\sigma_u = 5.$$

E x e m p l e 9. Pour localiser le défaut d'un ordinateur on effectue des tests. Par chacun des tests le défaut est localisé indépendamment des autres avec une probabilité $p = 0{,}2$. Un test dure environ trois minutes. Trouver l'espérance mathématique du temps nécessaire à la localisation du défaut.

S o l u t i o n. En utilisant le résultat du problème 9 du présent paragraphe (espérance mathématique du nombre d'expériences avant la k-ième réalisation de l'événement A), posant $k = 1$, on trouve le nombre moyen de tests:

$$m_x = \frac{1}{p} = \frac{1}{0{,}2} = 5.$$

La durée moyenne de ces tests est:

$$5 \cdot 3 = 15 \text{ (minutes)}.$$

E x e m p l e 10. Des coups sont tirés sur un réservoir d'essence. Chaque coup a une probabilité d'atteinte égale à 0,3. Les coups sont indépendants. Le premier coup atteignant le réservoir donne lieu seulement à une fuite, à la suite du second coup l'essence s'enflamme, après quoi le tir est arrêté. Trouver l'espérance mathématique du nombre de coups tirés.

S o l u t i o n. En utilisant la même formule que précédemment on peut trouver l'espérance mathématique du nombre de coups précédant la seconde atteinte:

$$m_x = \frac{2}{0{,}3} = 6{,}67.$$

E x e m p l e 11. La probabilité de détecter un objectif à l'aide d'un radar croît avec le nombre de cycles d'exploration suivant la loi:

$$P(n) = 1 - 0{,}8^n,$$

où n est le nombre de cycles à partir du début de l'observation.

Trouver l'espérance mathématique du nombre de cycles précédant la détection.

S o l u t i o n. En utilisant les résultats du problème 10 du présent paragraphe on a:

$$m_x = \sum_{k=0}^{\infty} [1 - P(k)] = \sum_{k=0}^{\infty} 0{,}8^n = \frac{1}{1 - 0{,}8} = 5.$$

E x e m p l e 12. Pour recueillir une certaine information sur les conditions à la surface d'une planète peu explorée, des stations d'observation sont lancées dans la région en question. Chaque station atteint la région avec une probabilité égale à 0,7. Pour recueillir l'information nécessaire il suffit trois stations. Une seule station ne peut pas s'acquitter de la tâche, deux stations le feront avec une probabilité égale à 0,4. La liaison avec la région est permanente; les stations supplémentaires ne sont lancées que si la tâche n'est pas accomplie. Trouver l'espérance mathématique du nombre de stations à lancer.

S o l u t i o n. Soit X le nombre de stations parvenues dans la région d'étude et suffisantes pour recueillir l'information nécessaire. Dans le problème 10 du présent paragraphe nous avons trouvé l'espérance mathématique du nombre d'expériences nécessaires pour obtenir un certain résultat dont la probabilité

croît avec le nombre d'expériences suivant la loi $P(n)$. Cette espérance mathématique est égale à:

$$m_x = \sum_{k=0}^{\infty} [1 - P(k)].$$

Dans le cas présent:

$$P(0) = 0; \quad P(1) = 0; \quad P(2) = 0{,}4; \quad P(3) = 1;$$
$$P(4) = P(5) = \ldots 1.$$

L'espérance mathématique de la variable aléatoire X est égale à:

$$m_x = M[X] = \sum_{k=0}^{\infty} [1 - P(k)] = 1 + 1 + 0{,}6 = 2{,}6.$$

Ainsi pour rassurer l'accomplissement de la tâche il faut que dans la région en question arrivent en moyenne 2,6 stations.

Proposons-nous de résoudre maintenant le problème suivant. Combien de stations faut-il envoyer en moyenne dans une région donnée pour que 2,6 stations arrivent à destination? Nous allons raisonner comme suit. Supposons qu'on envoie Y stations. Le nombre de stations arrivant dans la région en question peut s'écrire comme la somme:

$$X = Z_1 + Z_2 + \ldots + Z_y,$$

où $Z = 1$ si l'i-ème station arrive à destination et $Z = 0$ dans le cas contraire. La grandeur X n'est rien d'autre que la somme d'un nombre aléatoire de composantes aléatoires (voir problème 11 du présent paragraphe). On a ainsi

$$m_x = M[X] = M[Z_i]\, M[Y],$$

d'où:

$$M[Y] = \frac{M[X]}{M[Z_i]} = \frac{2{,}6}{M[Z_i]}.$$

Mais $M[Z_i] = p$ où p est la probabilité d'arrivée dans la région en question d'une station lancée (dans notre cas $p = 0{,}7$). On a:

$$m_y = M[Y] = \frac{2{,}6}{0{,}7} \approx 3{,}71.$$

Ainsi pour recueillir l'information dans une région donnée d'une planète il faut y envoyer en moyenne 3,7 stations.

Exemple 13. Une station radar explore une région où se trouvent N objectifs. La probabilité de détecter chacun des objectifs est égale à p (indépendamment de la détection des autres objectifs). Chaque cycle dure τ. Combien de temps faut-il pour détecter en moyenne k objectifs d'un ensemble de N.

Solution. Nous allons tout d'abord trouver l'espérance mathématique du nombre des objectifs détectés à la suite de n cycles d'exploration. Après n cycles un quelconque des objectifs est détecté avec une probabilité égale à:

$$P_n = 1 - (1 - p)^n;$$

en vertu du théorème d'addition des espérances mathématiques, le nombre moyen d'objectifs détectés dans n cycles est (voir problème 5 du présent paragraphe):

$$M[X] = N[1 - (1 - p)^n].$$

Posant

$$N[1 - (1 - p)^n] = k,$$

on trouve l'équation donnant le nombre nécessaire de cycles:

$$1 - (1 - p)^n = \frac{k}{N}.$$

La solution de cette équation est:

$$n=\frac{\log\left(1-\frac{k}{N}\right)}{\log(1-p)},$$

d'où le temps nécessaire pour détecter en moyenne k objectifs:

$$t_k=\tau n=\tau\,\frac{\log\left(1-\frac{k}{N}\right)}{\log(1-p)}.$$

Exemple 14. Modifions les conditions de l'exemple 13. Supposons que la station radar explore la région seulement jusqu'à ce que k objectifs soient détectés, après quoi l'exploration est interrompue ou poursuivie dans une autre région. Trouver l'espérance mathématique du temps nécessaire à cette opération.

Pour résoudre ce problème il ne suffit plus de connaître la probabilité de détection d'un objectif dans un cycle, il faut encore indiquer la manière dont augmente, avec le nombre de cycles, la probabilité pour que d'un groupe de N objectifs au moins k soient détectés. Pour calculer cette probabilité le plus simple est de supposer que les objectifs sont détectés indépendamment les uns des autres. Nous allons résoudre ce problème dans cette hypothèse.

Solution. Lorsque les détections sont indépendantes, détecter n objectifs équivaut à réaliser n expériences indépendantes. Après n cycles chacun des objectifs est détecté avec une probabilité égale à:

$$P_n=1-(1-p)^n.$$

Fig. 10.3.2

Le théorème des expériences répétées permet de trouver la probabilité pour que dans n cycles, k au moins d'un groupe de N objectifs soient détectés:

$$R_{k,N}^{(n)}=\sum_{m=k}^{N}C_N^m P_n^m(1-P_n)^{N-m}.$$

Le nombre moyen de cycles après lesquels k objectifs au moins seront détectés est donné par la formule (10.3.22):

$$n_{\text{moy}}^{(k)}=\sum_{n=0}^{\infty}[1-R_{k,N}^{(n)}]=\sum_{n=0}^{N}\left[1-\sum_{m=K}^{N}C_N^m P_n^m(1-P_n)^{N-m}\right].$$

Exemple 15. Le point aléatoire M du plan xOy de coordonnées (X, Y) s'écarte de la position requise (origine des coordonnées) sous l'influence de trois erreurs vectorielles $\vec{V}_1$, $\vec{V}_2$ et $\vec{V}_3$. Chacun des vecteurs (fig. 10.3.2) est caractérisé par deux composantes, soient:

$$\vec{V}_1\,(X_1, Y_1);\quad \vec{V}_2\,(X_2, Y_2);\quad \vec{V}_3\,(X_3, Y_3).$$

Les caractéristiques numériques de ces vecteurs sont:

$$m_{x_1}=2,\quad m_{y_1}=-3,\quad \sigma_{x_1}=2,$$
$$m_{x_2}=-1,\quad m_{y_2}=-2,\quad \sigma_{x_2}=4,$$
$$m_{x_3}=3,\quad m_{y_3}=1,\quad \sigma_{x_3}=2,$$
$$\sigma_{y_1}=3,\quad r_{x_1y_1}=-0,3;$$
$$\sigma_{y_2}=1,\quad r_{x_2y_2}=0,5;$$
$$\sigma_{y_3}=2,\quad r_{x_3y_3}=0,2.$$

Trouver les caractéristiques de l'erreur totale (vecteur de l'écart du point M de l'origine des coordonnées).

S o l u t i o n. En vertu du théorème d'addition des espérances mathématiques, des variances et des covariances on a:

$$m_x=m_{x_1}+m_{x_2}+m_{x_3}=4,$$
$$m_y=m_{y_1}+m_{y_2}+m_{y_3}=-4,$$
$$\sigma_x^2=\sigma_{x_1}^2+\sigma_{x_2}^2+\sigma_{x_3}^2=24,\quad \sigma_x=\sqrt{24}\approx 4,90,$$
$$\sigma_y^2=\sigma_{y_1}^2+\sigma_{y_2}^2+\sigma_{y_3}^2=14,\quad \sigma_y=\sqrt{14}\approx 3.75,$$
$$K_{xy}=K_{x_1y_1}+K_{x_2y_2}+K_{x_3y_3},$$

où

$$K_{x_1y_1}=r_{x_1y_1}\sigma_{x_1}\sigma_{y_1}=-0,3\cdot 2\cdot 3=-1,8,$$
$$K_{x_2y_2}=r_{x_2y_2}\sigma_{x_2}\sigma_{y_2}=0,5\cdot 4\cdot 1=2,0,$$
$$K_{x_3y_3}=r_{x_3y_3}\sigma_{x_3}\sigma_{y_3}=0,2\cdot 2\cdot 2=0,8,$$

d'où

$$K_{xy}=-1,8+2,0+0,8=1,0$$

et

$$r_{xy}=\frac{1,0}{4,90\cdot 3,75}\approx 0,054.$$

E x e m p l e 16. Un corps ayant la forme d'un parallélépipède rectangle de cotes a, b, c vole dans l'espace en tournant autour de son centre de masse de telle sorte que toutes ces orientations sont équiprobables. Le corps se meut dans un flux de particules et le nombre moyen de particules le rencontrant est proportionnel à l'aire moyenne de la surface se présentant au flux. Trouver l'espérance mathématique de l'aire de la projection du corps sur un plan perpendiculaire à la direction du mouvement.

S o l u t i o n. Toutes les orientations du corps dans l'espace étant équiprobables, la direction du plan de projection peut être quelconque. Il est évident que l'aire de la projection du corps est égale à la moitié de la somme des projections de toutes les faces du parallélépipède (car chaque point de la projection est la projection de deux points de la surface du corps). En vertu du théorème d'addition des espérances mathématiques et de la formule de l'aire moyenne de la projection d'une figure plane (voir exemple 3 du § 10.1) on a:

$$m_S=\frac{ab}{2}+\frac{ac}{2}+\frac{bc}{2}=\frac{S}{4},$$

où S est l'aire totale de la surface du parallélépipède

Notons que la formule obtenue est vraie non seulement pour un parallélépipède mais également pour tout corps convexe: l'aire moyenne de la projection d'un tel corps pour une rotation désordonnée est égale au quart de son aire totale. Nous recommandons au lecteur de démontrer cette assertion à titre d'exercice.

Exemple 17. Un point x se meut d'une manière aléatoire le long de l'axe des abscisses Ox. Au moment initial il se trouve à l'origine des coordonnées et commence à se mouvoir à droite avec une probabilité égale à 1/2 et à gauche avec une probabilité 1/2. Après avoir parcouru un segment unitaire le point continue son mouvement dans la même direction avec une probabilité égale à p ou change de sens avec une probabilité $q = 1 - p$. Après avoir parcouru un segment unitaire il peut de nouveau continuer son mouvement dans la même direction avec la probabilité p ou changer de sens avec la probabilité $1 - p$, etc.

Par suite de ce mouvement errant suivant l'axe des abscisses, après n pas le point x se trouvera à une distance aléatoire X_n de l'origine. Trouver l'espérance mathématique et la variance de X_n.

Solution. Par suite de la symétrie du problème, on a $M[X_n] = 0$. Pour trouver $D[X_n]$ nous allons écrire X_n comme une somme de n termes:

$$X_n = U_1 + U_2 + \ldots + U_n = \sum_{i=1}^{n} U_i, \qquad (10.3.31)$$

où U_i est la distance parcourue par le point pendant l'i-ème pas, c'est-à-dire $+1$ si le point se déplaçait vers la droite et -1 s'il se déplaçait vers la gauche.

En vertu du théorème sur la variance d'une somme [voir formule (10.2.10)] on a:

$$D[X_n] = \sum_{i=1}^{n} D[U_i] + 2 \sum_{i<j} K_{u_i u_j}.$$

Il est évident que $D[U_i] = 1$, car la variable U_i prend les valeurs $+1$ et -1 (en vertu des mêmes conditions de symétrie). Les covariances sont:

$$K_{u_i u_j} = M[\mathring{U}_i \mathring{U}_j] = M[U_i U_j].$$

On commence par le cas où $j = i + 1$ quand les variables U_i et U_j sont voisines dans la somme (10.3.31). Il est évident que $U_i U_{i+1}$ prend la valeur $+1$ avec la probabilité p et la valeur -1 avec la probabilité q. On a:

$$K_{u_i u_{i+1}} = M[U_i U_{i+1}] = 1 \cdot p + (-1) \cdot q = p - q.$$

Considérons ensuite le cas $j = i + 2$. Dans ce cas le produit $U_i U_j$ est égal à $+1$, si l'i-ème et l'$i + 2$-ème pas étaient dirigés dans le même sens. Ceci peut avoir lieu de deux manières différentes. Ou bien le point x fait tous les trois pas: l'i-ème, l'$i + 1$-ième et l'$i + 2$-ième dans la même direction, ou bien cette direction a changé deux fois durant ces trois pas. La probabilité pour que $U_i U_{i+2} = 1$ est égale à:

$$P(U_i U_{i+2} = 1) = P((U_i U_{i+1} = 1)(U_{i+1} U_{i+2} = 1)) + \\ + P((U_i U_{i+1} = -1)(U_{i+1} U_{i+2} = -1)) = p^2 + q^2.$$

Nous allons maintenant trouver la probabilité pour que $U_i U_{i+2} = -1$. Ceci peut également avoir lieu de deux manières différentes, le point peut changer de direction en passant du i-ème au $i + 1$-ième pas et conserver cette direc-

tion en passant du $i+1$-ième au $i+2$-ième pas, ou inversement. On a donc:

$$P(U_iU_{i+2}=1)=P((U_iU_{i+1}=-1)(U_{i+1}U_{i+2}=1))+$$
$$+P((U_iU_{i+1}=1)(U_{i+1}U_{i+2}=-1))=2pq.$$

Ainsi la variable U_iU_{i+2} prend deux valeurs possibles: $+1$ et -1, avec les probabilités respectives p^2+q^2 et $2pq$. Son espérance mathématique est égale à:

$$K_{u_iu_{i+2}}=M[U_iU_{i+2}]=p^2+q^2-2pq=(p-q)^2.$$

On peut facilement montrer par récurrence que pour un nombre quelconque k séparant les pas considérés dans la suite $U_1, U_2, \ldots, U_n$ on a les formules:

$$P(U_iU_{i+k}=1)=p^k+C_k^2p^{k-2}q^2+C_k^4p^{k-4}q^4+\ldots,$$
$$P(U_iU_{i+k}=-1)=C_k^1p^{k-1}q+C_k^3p^{k-3}q^3+\ldots,$$

par conséquent:

$$K_{u_iu_{i+k}}=(p-q)^k.$$

Ainsi la matrice de corrélation du système de variables aléatoires $U_1, U_2, \ldots$
$\ldots, U_n$ sera:

$$\|K_{u_iu_j}\|=\begin{Vmatrix}1 & p-q & (p-q)^2 & \ldots & (p-q)^{n-1}\\ & 1 & p-q & \ldots & (p-q)^{n-2}\\ & & 1 & \ldots & (p-q)^{n-3}\\ & & & & \cdots\cdots\\ & & & & 1\end{Vmatrix}.$$

La variance de la variable aléatoire X_n sera:

$$D[X_n]=\sum_{i=1}^{n}D[U_i]+2\sum_{i<j}K_{u_iu_j}=n+2\sum_{i<j}(p-q)_{j-i},$$

ou en prenant la somme des éléments se trouvant à une même distance de la diagonale principale:

$$D[X_n]=n+2\sum_{k=1}^{n-1}(n-k)(p-q)^k.$$

Exemple 18. Trouver l'asymétrie de la loi binomiale

$$P(X=m)=C_n^m p^m q^{n-m}\quad(q=1-p).\tag{10.3.32}$$

Solution. On sait que la loi binomiale (10.3.32) est la répartition du nombre d'apparitions dans n expériences indépendantes d'un certain événement qui a la probabilité p de se réaliser dans chacune des expériences. On écrit la variable aléatoire X, nombre d'apparitions de l'événement dans n experiences, comme la somme de n variables aléatoires.

$$X=\sum_{i=1}^{n}X_i,$$

où

$$x_i=\begin{cases}1 \text{ si dans l'}i\text{-ème expérience l'événement est apparu;}\\ 0 \text{ si dans l'}i\text{-ème expérience l'événement n'est pas apparu.}\end{cases}$$

En vertu du théorème d'addition des moments centrés trois on a:

$$\mu_3[X] = \sum_{i=1}^{n} \mu_3[X_i]. \tag{10.3.33}$$

. Nous allons calculer le moment centré trois de X_i. Sa loi de répartition est donnée par le tableau:

0	1
q	p

Le moment centré trois de X_i est égal à $(0-p)^3 q + (1-q)^3 p = -p^3 q + q^3 p = pq(q-p)$.
La substitution dans (10.3.33) donne:

$$\mu_3[X] = \sum_{i=1}^{n} pq(q-p) = npq(q-p).$$

Pour trouver l'asymétrie il faut diviser le moment centré trois de X par le cube de l'écart quadratique moyen:

$$Sk = \frac{npq(q-p)}{(npq)^{3/2}} = \frac{q-p}{\sqrt{npq}}.$$

E x e m p l e 19. Soient n variables aléatoires positives de même loi de répartition:

$$X_1, X_2, \ldots, X_n.$$

Trouver l'espérance mathématique de la variable aléatoire

$$Z_1 = \frac{X_1}{X_1 + X_2 + \ldots + X_n}.$$

S o l u t i o n. Il est clair que l'espérance mathématique de la variable Z_1 existe car elle se trouve entre zéro et l'unité. Il est de plus facile de voir que la loi de répartition du système $(X_1, X_2, \ldots, X_n)$, quelle qu'elle soit, est symétrique par rapport à ces variables, c'est-à-dire reste inchangée lors d'une permutation quelconque. Considérons les variables aléatoires

$$Z_1 = \frac{X_1}{X_1 + X_2 + \ldots + X_n}; \quad Z_2 = \frac{X_2}{X_1 + X_2 + \ldots + X_n}, \ldots$$

$$\ldots, Z_n = \frac{X_n}{X_1 + X_2 + \ldots + X_n}.$$

Il est évident que leur loi de répartition doit également être symétrique, c'est-à-dire qu'elle ne doit pas changer lorsque l'on remplace un argument par un autre. Par conséquent:

$$M[Z_1] = M[Z_2] = \ldots = M[Z_n].$$

De plus, on sait que la somme des variables aléatoires $Z_1, Z_2, \ldots, Z_n$ est égale à l'unité, par conséquent, en vertu du théorème d'addition des espérances mathématiques on a:

$$M[Z_1] + M[Z_2] + \ldots + M[Z_n] = M[1] = 1,$$

d'où

$$M[Z_1] = M[Z_2] = \cdots = M[Z_n] = \frac{1}{n}.$$

CHAPITRE 11

LINÉARISATION DES FONCTIONS

11.1. Méthode de linéarisation des fonctions d'arguments aléatoires

Dans le chapitre précédent nous avons étudié un appareil mathématique très commode de la théorie des probabilités, à savoir l'appareil des caractéristiques numériques. Dans de nombreux cas il permet de trouver les caractéristiques numériques des grandeurs aléatoires (et en premier lieu l'espérance mathématique et la variance) d'après les caractéristiques numériques des arguments, laissant de côté les lois de répartition. Ces méthodes de détermination directe des caractéristiques numériques sont applicables essentiellement aux fonctions *linéaires*.

Dans la pratique on rencontre souvent des cas où la fonction étudiée des variables aléatoires bien que non strictement linéaire, diffère peu d'une fonction linéaire et peut dans certains problèmes être remplacée par une fonction linéaire. Ceci provient de ce que, dans la majorité des applications pratiques, les variations aléatoires des grandeurs figurant dans ces fonctions apparaissent comme des erreurs insignifiantes, se superposant à la loi de base. Ces erreurs étant relativement peu importantes, les fonctions que l'on rencontre dans les problèmes, bien que non linéaires dans tout le domaine de variation des arguments, se trouvent être *presque linéaires* dans un domaine étroit de leurs variations aléatoires.

En effet, on sait que toute fonction continue dérivable, dans un domaine assez restreint de variation des arguments peut approximativement être remplacée par une fonction linéaire (on dit que cette fonction est linéarisée). L'erreur apparaissant alors est d'autant plus petite que sont étroites les limites de variation des arguments et que la fonction est voisine de la loi linéaire. Si le domaine des valeurs que peuvent prendre les arguments aléatoires est suffisamment petit pour que dans ce domaine la fonction puisse avec une précision suffisante pour les applications pratiques, être linéarisée, on peut appliquer à cette dernière l'appareil des caractéristiques numériques élaboré pour les fonctions linéaires. Connaissant les caractéristiques numériques des arguments on peut alors trouver les caractéristiques numériques des fonctions. Evidemment, on aura alors une solution

approchée du problème, mais dans la majorité des cas, on n'a pas besoin de la solution exacte.

Dans les applications pratiques où les facteurs aléatoires donnent lieu à des perturbations insignifiantes se superposant aux lois de base, la linéarisation est presque toujours possible.

Considérons, par exemple, le problème de balistique extérieure concernant le mouvement du centre de masse d'un obus. La distance X de vol d'un obus se détermine comme une certaine fonction des conditions de tir, à savoir de l'angle de jetée θ_0, de la vitesse initiale v_0 et du coefficient balistique c:

$$X = \varphi(\theta_0, v_0, c). \tag{11.1.1}$$

La fonction (11.1.1) est non linéaire dans tout le domaine de variation des arguments. C'est pourquoi dans le problème de base de balistique extérieure la fonction (11.1.1) est non linéaire et ne peut pas être linéarisée. Néanmoins, il y a des cas où ces fonctions peuvent être linéarisées; ce sont les problèmes liés à l'étude des *erreurs*. Supposons qu'il y ait lieu de trouver l'erreur aléatoire de la distance de vol X d'un obus, liée à la présence d'un certain nombre de facteurs aléatoires: l'imprécision sur l'angle θ_0, les oscillations du tube lors du coup, la non-homogénéité balistique des obus, les poids différents des charges, etc. Nous pouvons alors fixer certaines conditions nominales du tir et considérer les écarts accidentels à ces conditions. La gamme de ces variations aléatoires est, en général, restreinte et la fonction φ, bien que non linéaire dans tout le domaine de ses arguments, peut être linéarisée dans un domaine étroit de leurs variations aléatoires.

La méthode de linéarisation des fonctions des arguments aléatoires est couramment appliquée dans de nombreux domaines techniques. Très souvent, après avoir obtenu la solution d'un problème par les méthodes ordinaires des sciences exactes, on désire estimer les erreurs possibles de cette solution dues à l'influence de facteurs aléatoires dont on n'a pas tenu compte. Dans ce cas le problème peut être résolu par linéarisation, car les variations aléatoires figurant dans le problème ne sont, en général, pas très importantes. S'il n'en était pas ainsi et si les variations aléatoires des arguments se trouvaient au-delà du domaine de linéarisation approchée des fonctions, la réalisation technique envisagée ne serait pas satisfaisante car son indétermination serait trop grande.

11.2. Linéarisation d'une fonction d'un argument aléatoire

Dans la pratique il est rare qu'il y ait lieu de linéariser une fonction d'un seul argument aléatoire, en général, on est obligé de tenir compte de l'influence d'un ensemble de facteurs aléatoires. Cepen-

dant, du point de vue méthodique, il est commode de commencer par ce cas le plus simple. Soit la variable aléatoire X dont on connaît les caractéristiques numériques : l'espérance mathématique m_x et la variance D_x.

Supposons que les valeurs possibles de X se situent entre α et β, c'est-à-dire :

$$P(\alpha < X < \beta) \approx 1.$$

Soit maintenant une autre variable aléatoire Y, liée à X par la relation fonctionnelle

$$Y = \varphi(X)\ ^{*}), \qquad (11.2.1)$$

la fonction φ, bien que non linéaire, diffère peu d'une loi linéaire sur le segment (α, β).

Il y a lieu de trouver les caractéristiques numériques de Y, à savoir son espérance mathématique m_y et sa variance D_y.

Considérons la courbe $y = \varphi(x)$ dans l'intervalle (α, β) (fig. 11.2.1) et remplaçons-la par la tangente au point M d'abscisse m_x. Cette tangente a pour équation :

$$y = \varphi(m_x) + \varphi'(m_x)(x - m_x). \qquad (11.2.2)$$

Fig. 11.2.1

Supposons que l'intervalle des valeurs possibles de l'argument (α, β) soit suffisamment étroit pour que la courbe et la tangente diffèrent peu. On peut alors remplacer la courbe pratiquement par un segment de tangente ; dans l'intervalle (α, β) la fonction $y = \varphi(x)$ se trouve être presque linéaire. Les variables aléatoires X et Y sont approximativement liées par la relation linéaire :

$$Y = \varphi(m_x) + \varphi'(m_x)(X - m_x),$$

ou, en introduisant la désignation $X - m_x = \overset{\circ}{X}$:

$$Y = \varphi(m_x) + \varphi'(m_x)\overset{\circ}{X}. \qquad (11.2.3)$$

On peut appliquer à la fonction linéaire (11.2.3) les méthodes connues de recherche des caractéristiques numériques des fonctions linéaires (voir § 10.2). Pour trouver l'espérance mathématique de cette fonction linéaire il y a lieu de substituer dans son expression (11.2.3) l'espérance mathématique de $\overset{\circ}{X}$ égale à zéro. On obtient :

$$m_y = \varphi(m_x). \qquad (11.2.4)$$

*) On suppose que sur le segment (α, β) la fonction φ est continue et dérivable.

La variance de Y est donnée par la formule:

$$D_y = [\varphi'(m_x)]^2 D_x. \qquad (11.2.5)$$

En passant à l'écart quadratique moyen, on a:

$$\sigma_y = |\varphi'(m_x)| \sigma_x. \qquad (11.2.6)$$

Les formules (11.2.4), (11.2.5), (11.2.6) sont évidemment approchées car le fait de remplacer une fonction non linéaire par une fonction linéaire est déjà une approximation.

Nous avons ainsi résolu le problème posé et nous sommes arrivés aux conclusions suivantes.

Pour trouver l'espérance mathématique d'une fonction presque linéaire il faut remplacer, dans l'expression de la fonction, l'argument par son espérance mathématique. Pour trouver la variance d'une fonction presque linéaire il faut multiplier la variance de l'argument par le carré de la dérivée au point correspondant à l'espérance mathématique donnée de l'argument.

11.3. Linéarisation d'une fonction de plusieurs arguments aléatoires

Soit un système de n variables aléatoires:

$$(X_1, X_2, \ldots, X_n)$$

et soient données les caractéristiques numériques du système, à savoir les espérances mathématiques:

$$m_{x_1}, m_{x_2}, \ldots, m_{x_n}$$

et la matrice de corrélation:

$$\| K_{ij} \| = \left\| \begin{matrix} K_{11} & K_{12} & \ldots & K_{1n} \\ & K_{22} & \ldots & K_{2n} \\ & & \ldots & \ldots \\ & & & K_{nn} \end{matrix} \right\| .$$

La variable aléatoire Y est une fonction des arguments X_1, $X_2, \ldots, X_n$:

$$Y = \varphi(X_1, X_2, \ldots, X_n). \qquad (11.3.1)$$

La fonction φ n'est pas linéaire, mais diffère peu d'une fonction linéaire dans le domaine des valeurs possibles de tous les arguments, c'est-à-dire, c'est une fonction presque linéaire. Il y a lieu de trouver les caractéristiques numériques de Y, donc l'espérance mathématique m_y et la variance D_y.

Pour résoudre ce problème il y a lieu de linéariser la fonction:

$$y = \varphi(x_1, x_2, \ldots, x_n). \qquad (11.3.2)$$

On ne peut pas donner ici l'interprétation géométrique car au-delà de l'espace tridimensionnel elle n'est plus figurative. Cependant qualitativement tout est analogue au cas étudié au paragraphe précédent.

Considérons la fonction $y = \varphi(x_1, x_2, \ldots, x_n)$ au voisinage immédiat du point $m_{x_1}, m_{x_2}, \ldots, m_{x_n}$. La fonction au voisinage de ce point étant presque linéaire, elle peut être approximativement remplacée par une fonction linéaire. Ceci équivaut à ne conserver dans son développement en série de Taylor au voisinage du point $(m_{x_1}, m_{x_2}, \ldots, m_{x_n})$ que les termes du premier ordre et à négliger les autres termes:

$$y = \varphi(x_1, x_2, \ldots, x_n) \approx \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \\ + \sum_{i=1}^{n} \varphi'_{x_i}(m_{x_1}, m_{x_2}, \ldots, m_{x_n})(x_i - m_{x_i}).$$

Ceci signifie que la relation (11.3.1) entre les variables aléatoires peut être remplacée par une relation linéaire:

$$Y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \sum_{i=1}^{n} \varphi'_{x_i}(m_{x_1}, m_{x_2}, \ldots, m_{x_n})(X_i - m_{x_i}). \tag{11.3.3}$$

Pour simplifier introduisons les désignations:

$$\varphi'_{x_i}(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) = \left(\frac{\partial \varphi}{\partial x_i}\right)_m.$$

Comme $X_i - m_{x_i} = \overset{\circ}{X}_i$ la formule (11.3.3) s'écrit comme suit:

$$Y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \sum_{i=1}^{n} \left(\frac{\partial \varphi}{\partial x_i}\right)_m \overset{\circ}{X}_i. \tag{11.3.4}$$

On peut appliquer à la fonction linéaire (11.3.4) les méthodes de calcul des caractéristiques numériques des fonctions linéaires établies au § 10.2. Les arguments centrés $(\overset{\circ}{X}_1, \overset{\circ}{X}_2, \ldots, \overset{\circ}{X}_n)$ ayant des espérances mathématiques nulles et la même matrice de corrélation $\| K_{ij} \|$, on a:

$$m_y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}), \tag{11.3.5}$$

$$D_y = \sum_{i=1}^{n} \left(\frac{\partial \varphi}{\partial x_i}\right)_m^2 D_{x_i} + 2 \sum_{i<j} \left(\frac{\partial \varphi}{\partial x_i}\right)_m \left(\frac{\partial \varphi}{\partial x_j}\right)_m K_{ij}. \tag{11.3.6}$$

En passant dans la dernière formule des variances aux écarts quadratiques moyens on obtient:

$$\sigma_y^2 = \sum_{i=1}^{n} \left(\frac{\partial \varphi}{\partial x_i}\right)_m^2 \sigma_{x_i}^2 + 2 \sum_{i<j} \left(\frac{\partial \varphi}{\partial x_i}\right)_m \left(\frac{\partial \varphi}{\partial x_j}\right)_m r_{ij} \sigma_{x_i} \sigma_{x_j}, \tag{11.3.7}$$

où r_{ij} est le coefficient de corrélation des grandeurs X_i, X_j.

La formule (11.3.7) est particulièrement simple lorsque les variables $X_1, X_2, \ldots, X_n$ ne sont pas corrélées, c'est-à-dire lorsque $r_{ij} = 0$ pour $i \neq j$. Dans ce cas:

$$\sigma_y^2 = \sum_{i=1}^{n} \left(\frac{\partial \varphi}{\partial x_i}\right)_m^2 \sigma_{x_i}^2. \qquad (11.3.8)$$

Les formules (11.3.7) et (11.3.8) ont de nombreuses applications pratiques, par exemple, l'étude des erreurs des différents appareils et mécanismes ou la précision des tirs et des bombardements.

11.4. Précision des résultats obtenus par la méthode de linéarisation

Dans certains problèmes pratiques la gamme des variations possibles des arguments aléatoires n'est pas suffisamment petite pour que, dans ses limites, la fonction puisse être linéarisée avec une précision suffisante.

Il est conseillé dans ce cas, pour juger de l'applicabilité de la méthode de linéarisation et pour préciser les résultats obtenus, d'utiliser la méthode suivant laquelle on conserve dans le développement de la fonction, outre les termes linéaires, plusieurs termes d'ordre supérieur, en estimant toujours l'erreur due à ces termes.

Pour préciser cette méthode nous allons considérer tout d'abord le cas simple d'un seul argument. La variable aléatoire Y est une fonction de l'argument aléatoire X:

$$Y = \varphi(X), \qquad (11.4.1)$$

la fonction φ, étant presque linéaire dans le domaine des valeurs possibles de l'argument X, présente des écarts suffisamment grands pour que l'on puisse mettre en doute l'applicabilité de la méthode de linéarisation. Pour le vérifier nous allons appliquer une méthode plus précise, à savoir le développement de φ en série de Taylor au voisinage du point m_x, en ne conservant dans le développement que les trois premiers termes:

$$y = \varphi(x) \approx \varphi(m_x) + \varphi'(m_x)(x - m_x) + \frac{1}{2}\varphi''(m_x)(x - m_x)^2. \qquad (11.4.2)$$

Les variables aléatoires Y et X sont évidemment liées par la même relation:

$$Y = \varphi(m_x) + \varphi'(m_x)(X - m_x) + \frac{1}{2}\varphi''(m_x)(X - m_x)^2 =$$

$$= \varphi(m_x) + \varphi'(m_x)\overset{\circ}{X} + \frac{1}{2}\varphi''(m_x)\overset{\circ}{X}{}^2. \qquad (11.4.3)$$

En utilisant l'expression (11.4.3) on peut trouver l'espérance mathématique et la variance de Y. En appliquant les théorèmes sur

les caractéristiques numériques on a :

$$m_y = \varphi(m_x) + \frac{1}{2}\varphi''(m_x) M[\overset{\circ}{X}{}^2] = \varphi(m_x) + \frac{1}{2}\varphi''(m_x) D_x. \qquad (11.4.4)$$

La formule (11.4.4) donne la valeur précisée de l'espérance mathématique qu'on peut comparer à la valeur de $\varphi(m_x)$ obtenue par linéarisation ; la correction tenant compte de la non-linéarité de la fonction est donnée par le second terme de la formule (11.4.4).

Le calcul de la variance des premier et second membres de la formule (11.4.3) donne :

$$D_y = [\varphi'(m_x)]^2 D_x + \frac{1}{4}[\varphi''(m_x)]^2 D[\overset{\circ}{X}{}^2] + $$
$$+ \varphi'(m_x)\varphi''(m_x) K[\overset{\circ}{X}, \overset{\circ}{X}{}^2], \qquad (11.4.5)$$

où $K[\overset{\circ}{X}, \overset{\circ}{X}{}^2]$ est la covariance des valeurs $\overset{\circ}{X}, \overset{\circ}{X}{}^2$.

Exprimons les grandeurs entrant dans la formule (11.4.5) en fonction des moments centrés de X :

$$D[\overset{\circ}{X}{}^2] = M[\overset{\circ}{X}{}^4] - \{M[\overset{\circ}{X}{}^2]\}^2 = \mu_4[X] - D_x^2,$$
$$K[\overset{\circ}{X}, \overset{\circ}{X}{}^2] = M[\overset{\circ}{X}\{X^2 - M[\overset{\circ}{X}{}^2]\}] = \mu_3[X].$$

Finalement on obtient :

$$D_y = [\varphi'(m_x)]^2 D_x + \frac{1}{4}[\varphi''(m_x)]^2 (\mu_4[X] - D_x^2) +$$
$$+ \varphi'(m_x)\varphi''(m_x)\mu_3[X]. \qquad (11.4.6)$$

La formule (11.4.6) donne une valeur plus précise de la variance que la méthode de linéarisation ; les second et troisième termes donnent la correction due à la non-linéarité de la fonction. En plus de la variance D_x de l'argument la formule contient également les moments centrés trois et quatre $\mu_3[X]$, $\mu_4[X]$. Lorsque l'on connaît ces moments la correction à la variance est donnée directement par la formule (11.4.6). Cependant, souvent il n'est pas nécessaire de la déterminer, il suffit d'en connaître l'ordre de grandeur. Dans la pratique on rencontre souvent des variables aléatoires obéissant à la loi normale. Pour une variable aléatoire normale on a :

$$\mu_3[X] = 0 ; \quad \mu_4[X] = 3\sigma_x^4 = 3D_x^2 \qquad (11.4.7)$$

et la formule (11.4.6) devient alors :

$$D_y = [\varphi'(m_x)]^2 D_x + \frac{1}{2}[\varphi''(m_x)]^2 D_x^2. \qquad (11.4.8)$$

On peut utiliser la formule (11.4.8) pour estimer approximativement l'erreur de linéarisation lorsque l'argument est réparti suivant une loi voisine de la loi normale.

On peut appliquer une méthode analogue à une fonction de plusieurs arguments aléatoires:

$$Y = \varphi(X_1, X_2, \ldots, X_n). \qquad (11.4.9)$$

En développant la fonction

$$y = \varphi(x_1, x_2, \ldots, x_n)$$

en série de Taylor au voisinage du point $m_{x_1}, m_{x_2}, \ldots, m_{x_n}$ et en ne conservant que les termes d'ordre non supérieur à deux, on obtient l'approximation suivante:

$$Y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \sum_{i=1}^{n} \left(\frac{\partial\varphi}{\partial x_i}\right)_m (X_i - m_{x_i}) +$$

$$+ \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m (X_i - m_{x_i})^2 + \sum_{i<j} \left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}\right)_m (X_i - m_{x_i})(X_j - m_{x_j})$$

ou, en introduisant les grandeurs mentionnées centrées:

$$Y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \sum_{i=1}^{n} \left(\frac{\partial\varphi}{\partial x_i}\right)_m \mathring{X}_i +$$

$$+ \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m \mathring{X}_i^2 + \sum_{i<j} \left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}\right)_m \mathring{X}_i\mathring{X}_j, \qquad (11.4.10)$$

où l'indice m signifie comme précédemment que dans la dérivée partielle les arguments X_i ont été remplacés par leurs espérances mathématiques m_{x_i}.

En appliquant à la formule (11.4.10) l'opération d'espérance mathématique on a:

$$m_y = \varphi(m_{x_1}, m_{x_2}, \ldots m_{x_n}) + \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m D_{x_i} +$$

$$+ \sum_{i<j} \left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}\right)_m K_{ij}, \qquad (11.4.11)$$

où K_{ij} est la covariance des valeurs X_i, X_j.

Dans le cas le plus important pour les applications pratiques lorsque les arguments $X_1, X_2, \ldots, X_n$ ne sont pas corrélés, la formule (11.4.11) devient:

$$m_y = \varphi(m_{x_1}, m_{x_2}, \ldots, m_{x_n}) + \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m D_{x_i}. \qquad (11.4.12)$$

Le second terme dans la formule (11.4.12) donne la correction due à la non-linéarité de la fonction.

Passons au calcul de la variable Y. Pour trouver son expression sous la forme la plus simple, supposons que les variables $X_1, X_2, \ldots$ $\ldots, X_n$ soient non seulement non corrélées mais également indépendantes. En calculant la variance des premier et second membres de (11.4.10) et en utilisant le théorème de la variance d'un produit (voir § 10.2) on a :

$$D_y = \sum_{i=1}^{n} \left(\frac{\partial\varphi}{\partial x_i}\right)_m^2 D_{x_i} + \frac{1}{4}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m^2 \cdot (\mu_4[X_i] - D_{x_i}^2) +$$

$$+ \sum_{i<j} \left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}\right)_m^2 D_{x_i} D_{x_j} + \sum \left(\frac{\partial\varphi}{\partial x_i}\right)_m \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m \mu_3[X_i]. \qquad (11.4.13)$$

Pour des variables réparties suivant une loi voisine de la loi normale on peut utiliser la formule (11.4.7) et écrire l'expression (11.4.13) comme suit :

$$D_y = \sum_{i=1}^{n} \left(\frac{\partial\varphi}{\partial x_i}\right)_m^2 D_{x_i} + \frac{1}{2}\sum_{i=1}^{n} \left(\frac{\partial^2\varphi}{\partial x_i^2}\right)_m^2 D_{x_i}^2 +$$

$$+ \sum_{i<j} \left(\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}\right)^2 D_{x_i} D_{x_j}. \qquad (11.4.14)$$

Les deux derniers termes dans l'expression (11.4.14) sont la correction due à la non-linéarité de la fonction et peuvent être utilisés pour estimer la précision de la méthode de linéarisation lors du calcul de la variance.

CHAPITRE 12

LOIS DE RÉPARTITION DES FONCTIONS DES ARGUMENTS ALÉATOIRES

12.1. Loi de répartition d'une fonction monotone d'un argument aléatoire

Dans les chapitres précédents nous avons étudié les méthodes de calcul des caractéristiques numériques des variables aléatoires; l'avantage essentiel de ces méthodes est de ne pas nécessiter la détermination des lois de répartition. Cependant parfois il est nécessaire de connaître non seulement les caractéristiques numériques mais également les lois de répartition des fonctions.

Nous allons commencer par le problème le plus simple de ce type, à savoir le problème de la détermination de la loi de répartition d'une fonction d'un seul argument aléatoire. Les grandeurs aléatoires continues se rencontrant le plus souvent dans les applications pratiques, nous allons commencer par elles.

Soit une variable aléatoire continue X de densité de probabilité $f(x)$ et une autre variable aléatoire Y fonctionnellement liée à X par la relation

$$Y = \varphi(X)\ ^{*}).$$

Il y a lieu de trouver la densité de probabilité de Y.

Considérons le segment (a, b) de l'axe des abscisses auquel appartiennent toutes les valeurs possibles de X, c'est-à-dire

$$P(a < X < b) = 1.$$

Dans le cas particulier où le domaine des valeurs possibles de X n'est pas limité on a $a = -\infty$; $b = +\infty$.

La méthode de résolution de ce problème dépend du comportement de la fonction φ sur le segment (a, b) : accroissement, décroissement ou oscillation.

Dans ce paragraphe nous allons envisager le cas où la fonction $y = \varphi(x)$ sur l'intervalle (a, b) est monotone **). Nous allons séparément étudier deux cas: fonction monotone croissante et fonction monotone décroissante.

---

*) La fonction φ est supposée continue et dérivable.

**) Le cas d'une fonction non monotone sera étudié dans le § 12.3.

1. La fonction $y = \varphi(x)$ sur l'intervalle (a, b) est monotone croissante (fig. 12.1.1.). Lorsque X prend différentes valeurs sur l'intervalle (a, b), le point aléatoire (X, Y) ne se déplace que sur la courbe $y = \varphi(x)$; l'ordonnée de ce point aléatoire est complètement déterminée par son abscisse.

Désignons par $g(y)$ la densité de probabilité de Y. Pour trouver $g(y)$, nous allons tout d'abord trouver la fonction de répartition de Y :

$$G(y) = P(Y < y).$$

Traçons la droite AB parallèlement à l'axe des abscisses à une distance y de celle-ci (fig. 12.1.1). Pour que la condition $Y < y$ puisse être remplie, il faut que le point aléatoire (X, Y) se situe sur

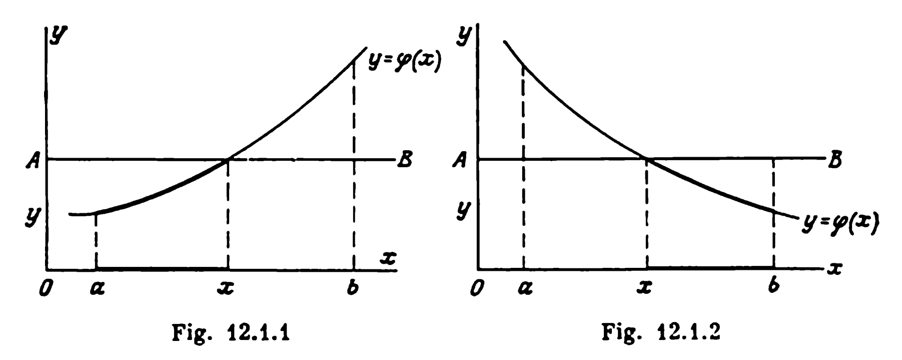


Fig. 12.1.1 Fig. 12.1.2

la portion de la courbe se trouvant au-dessous de la droite AB ; pour cela il faut et il suffit que la variable aléatoire X se trouve entre a et x, où x est l'abscisse du point d'intersection de la courbe $y = \varphi(x)$ avec la droite AB. Par conséquent :

$$G(y) = P(Y < y) = P(a < X < x) = \int_a^x f(x)\,dx.$$

La limite supérieure x de l'intégrale peut s'exprimer en fonction de y :

$$x = \psi(y),$$

où ψ est la fonction inverse de φ. On a alors :

$$G(y) = \int_a^{\psi(y)} f(x)\,dx. \qquad (12.1.1)$$

En dérivant l'intégrale (12.1.1) par rapport à la variable y figurant dans cette limite, on obtient :

$$g(y) = G'(y) = f(\psi(y))\,\psi'(y). \qquad (12.1.2)$$

2. La fonction $y = \varphi(x)$ dans l'intervalle (a, b) décroît d'une manière monotone (fig. 12.1.2). Dans ce cas :

$$G(y) = P(Y < y) = P(x < X < b) = \int_x^b f(x)\,dx = \int_{\psi(y)}^b f(x)\,dx,$$

d'où :

$$g(y) = G'(y) = -f(\psi(y))\,\psi'(y). \qquad (12.1.3)$$

En comparant les formules (12.1.2) et (12.1.3) on voit qu'elles peuvent être réunies en une seule :

$$g(y) = f(\psi(y))\,|\psi'(y)|. \qquad (12.1.4)$$

En effet, lorsque φ croît, sa dérivée (donc également ψ') est positive. Lorsque la fonction φ est décroissante la dérivée ψ' est négative, par contre devant cette fonction on a dans la formule (12.1.3) le signe moins. Par conséquent, la formule (12.1.4) dans laquelle la dérivée est prise en module est vraie dans les deux cas. Ainsi le problème de la recherche de la loi de répartition d'une fonction monotone se trouve résolu.

E x e m p l e. Soit x une variable aléatoire suivant une loi de Cauchy de densité de probabilité

$$f(x) = \frac{1}{\pi(1+x^2)}\ .$$

Les variables X et Y sont liées par la relation:

$$Y = 1 - X^3.$$

Trouver la densité de probabilité de Y.

S o l u t i o n. Comme la fonction $y = 1 - x^3$ est monotone sur l'intervalle $(-\infty, +\infty)$ on peut appliquer la formule (12.1.4). Nous allons représenter la solution de ce problème sous la forme de deux tableaux, dans celui de gauche nous écrivons les fonctions figurant dans la solution générale du problème, dans celui de droite, les fonctions correspondant à l'exemple étudié:

$f(x)$	$\dfrac{1}{\pi(1+x^2)}$
$y=\varphi(x)$	$y=1-x^3$
$x=\psi(y)$	$x=\sqrt[3]{1-y}$
$\psi'(y)$	$\dfrac{-1}{3\sqrt[3]{(1-y)^2}}$
$\lvert\psi'(y)\rvert$	$\dfrac{1}{3\sqrt[3]{(1-y)^2}}$
$g(y)=f(\psi(y))\lvert\psi'(y)\rvert$	$g(y)=\dfrac{1}{3\pi\,(1+\sqrt[3]{(1-y)^2})\,\sqrt[3]{(1-y)^2}}$

12.2. Loi de répartition d'une fonction linéaire d'un argument réparti suivant la loi normale

Soient X une variable normale de densité

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \qquad (12.2.1)$$

et Y une variable aléatoire liée à X par la relation fonctionnelle linéaire :

$$Y=aX+b, \qquad (12.2.2)$$

où a et b sont des coefficients non aléatoires.

Il y a lieu de trouver la loi de répartition de Y.

Nous allons écrire la solution sous la forme de deux tableaux comme dans l'exemple du paragraphe précédent :

$f(x)$	$\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}$
$y=\varphi(x)$	$y=ax+b$
$x=\psi(y)$	$x=\frac{y-b}{a}$
$\psi'(y)$	$\frac{1}{a}$
$\lvert\psi'(y)\rvert$	$\frac{1}{\lvert a\rvert}$
$g(y)=f(\psi(y))\lvert\psi'(y)\rvert$	$g(y)=\frac{1}{\lvert a\rvert\sigma_x\sqrt{2\pi}}e^{-\frac{\left(\frac{y-b}{a}-m_x\right)^2}{2\sigma_x^2}}$.

Après transformation $g(y)$ devient :

$$g(y)=\frac{1}{\lvert a\rvert\sigma_x\sqrt{2\pi}}e^{-\frac{[y-(am_x+b)]^2}{2\lvert a\rvert^2\sigma_x^2}}.$$

Cette expression n'est rien d'autre qu'une loi normale de paramètres :

$$\left.\begin{aligned} a_y &= am_x+b,\\ \sigma_y &= \lvert a\rvert\,\sigma_x. \end{aligned}\right\} \qquad (12.2.3)$$

En passant des écarts quadratiques moyens aux grandeurs proportionnelles que sont les écarts probables on a :

$$E_y=\lvert a\rvert E_x. \qquad (12.2.4)$$

On voit ainsi qu'*une fonction linéaire d'un argument aléatoire suivant une loi normale, suit également une loi normale*. Pour trouver

le centre de dispersion de cette loi, il faut dans l'expression de la fonction linéaire substituer à l'argument le centre de dispersion. Pour trouver l'écart quadratique moyen de cette loi il faut multiplier l'écart quadratique moyen de l'argument par le module du coefficient de l'argument dans l'expression de la fonction linéaire. Pour les écarts probables la règle est la même.

12.3. Loi de répartition d'une fonction non monotone d'un argument aléatoire

Soient X une variable aléatoire continue de densité de probabilité $f(x)$ et Y une autre variable liée à X par la relation fonctionnelle:

$$Y = \varphi(X),$$

la fonction $y = \varphi(x)$ n'étant pas monotone sur l'intervalle (a, b) des valeurs possibles de l'argument (fig. 12.3.1).

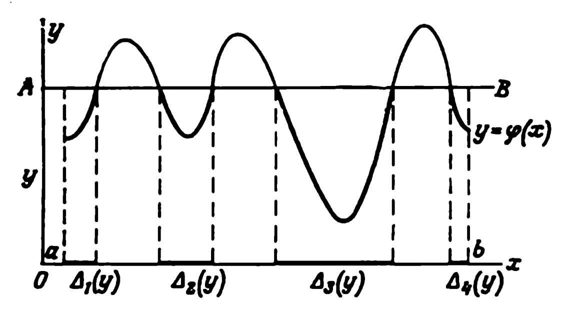


Fig. 12.3.1

Calculons la fonction de répartition $G(y)$ de Y. Pour cela traçons de nouveau une droite AB parallèle à l'axe des abscisses à une distance y de celui-ci et considérons les portions de la courbe $y = \varphi(x)$ où on a $Y < y$. Désignons par $\Delta_1(y)$, $\Delta_2(y)$ les intervalles correspondants de l'axe des abscisses.

L'événement $Y < y$ équivaut à ce que la variable aléatoire X se trouve dans l'un des intervalles $\Delta_1(y)$, $\Delta_2(y)$, ... Ainsi:

$$G(y) = P(Y < y) = P((X \subset \Delta_1(y)) + (X \subset \Delta_2(y)) + \ldots) =$$
$$= \sum_i P(X \subset \Delta_i(y)) = \sum_i \int_{\Delta_i(y)} f(x)\,dx.$$

La fonction de répartition de la grandeur $Y = \varphi(X)$ est donc donnée par la formule:

$$G(y) = \sum_i \int_{\Delta_i(y)} f(x)\,dx. \qquad (12.3.1)$$

Les limites des intervalles $\Delta_i(y)$ dépendent de y et pour une forme donnée de la fonction $y = \varphi(x)$ peuvent s'exprimer comme des fonctions explicites de y. En dérivant $G(y)$ par rapport à y figurant dans les bornes des intégrales, on obtient la densité de probabilité de Y:

$$g(y) = G'(y). \tag{12.3.2}$$

E x e m p l e. Soit une variable X équirépartie sur l'intervalle de $-\frac{\pi}{2}$ à $\frac{\pi}{2}$:

$$f(x) = \begin{cases} \frac{1}{\pi} & \text{pour } |x| < \frac{\pi}{2}; \\ 0 & \text{pour } |x| > \frac{\pi}{2}. \end{cases}$$

Trouver la loi de répartition de la grandeur

$$Y = \cos X.$$

S o l u t i o n. Traçons la courbe $y = \cos x$ (fig. 12.3.2). De toute évidence on a $a = -\frac{\pi}{2}$, $b = \frac{\pi}{2}$ et sur l'intervalle (a, b) la fonction $y = \cos x$ n'est pas monotone.

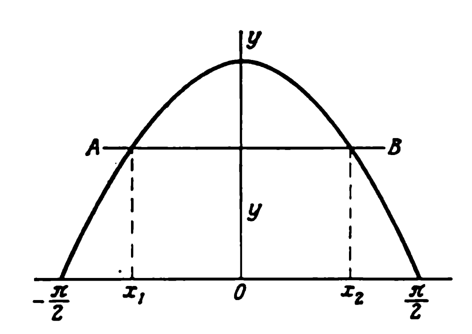


Fig. 12.3.2

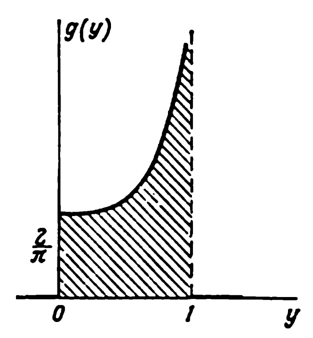


Fig. 12.3.3

En appliquant la formule (12.3.1) on a:

$$G(y) = \int_{-\frac{\pi}{2}}^{x_1} f(x)\,dx + \int_{x_2}^{\frac{\pi}{2}} f(x)\,dx.$$

Exprimons x_1 et x_2 en fonction de y:

$$x_1 = -\arccos y; \quad x_2 = \arccos y,$$

on obtient:

$$G(y) = \int_{-\frac{\pi}{2}}^{-\arccos y} f(x)\,dx + \int_{\arccos y}^{\frac{\pi}{2}} f(x)\,dx. \tag{12.3.3}$$

Pour trouver la densité $g(y)$ nous n'allons pas calculer les intégrales dans la formule (12.3.3) mais dérivons directement cette expression par rapport à la variable y entrant dans les bornes des intégrales:

$$g(y) = G'(y) = f(-\arccos y)\frac{1}{\sqrt{1-y^2}} + f(\arccos y)\frac{1}{\sqrt{1-y^2}}.$$

Comme $f(x) = \frac{1}{\pi}$ on a:

$$g(y) = \frac{2}{\pi}\frac{1}{\sqrt{1-y^2}}. \qquad (12.3.4)$$

Notons que la loi de répartition (12.3.4) pour Y n'est vraie qu'entre 0 et 1, c'est-à-dire dans les limites de variation de $Y = \cos X$ quand X varie entre $-\frac{\pi}{2}$ et $\frac{\pi}{2}$. Hors de ces limites la densité $g(y)$ est égale à zéro.

Le graphique de la fonction $g(y)$ est donné sur la figure 12.3.3. Pour $y = 1$, $g(y)$ tend vers l'infini.

12.4. Loi de répartition d'une fonction de deux variables aléatoires

Le calcul de la loi de répartition dans le cas d'une fonction de plusieurs variables est bien plus compliqué que dans celui d'une fonction d'un seul argument. Nous allons exposer ici la méthode générale permettant de résoudre ce problème pour le cas le plus simple de deux arguments.

Considérons un système de deux variables aléatoires continues (X, Y) de densité de probabilité $f(x, y)$. La variable aléatoire Z est liée à X et Y par la relation fonctionnelle:

$$Z = \varphi(X, Y).$$

Il y a lieu de trouver la loi de répartition de Z.

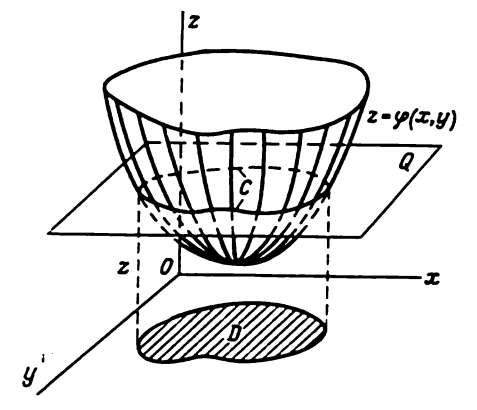


Fig. 12.4.1

Pour résoudre ce problème nous allons faire appel à une interprétation géométrique analogue à celle que nous avons utilisée dans le cas d'un seul argument. La fonction $z = \varphi(x, y)$ sera maintenant représentée non pas par une courbe mais par une surface (fig. 12.4.1).

Calculons la fonction de répartition de Z:

$$G(z) = P(Z < z) = P(\varphi(X, Y) < z). \qquad (12.4.1)$$

Soit le plan Q parallèle au plan xOy situé à une distance z de celui-ci. Ce plan coupera la surface $z = \varphi(x, y)$ suivant une certaine cour-

be C *). Projetons la courbe C sur le plan xOy. Cette projection dont l'équation est $\varphi(x, y) = z$ divisera le plan xOy en deux domaines: pour l'un de ces domaines la cote de la surface examinée sera inférieure et pour l'autre supérieure à z. Désignons par D le domaine où cette hauteur est inférieure à z. Pour que l'inégalité (12 4.1) soit vérifiée, il faut que le point aléatoire (X, Y) se trouve dans le domaine D ; par conséquent :

$$G(z) = P((X, Y) \subset D) = \iint_{(D)} f(x, y)\, dx\, dy. \qquad (12.4.2)$$

Dans l'expression (12.4.2) la grandeur z entre implicitement par l'intermédiaire des limites.

En dérivant $G(z)$ par rapport à z on obtient la densité de probabilité de Z:

$$g(z) = G'(z).$$

Connaissant la forme de la fonction $z = \varphi(x, y)$ on peut exprimer les limites d'intégration en fonction de z et expliciter $g(z)$.

Pour trouver la loi de répartition d'une fonction de deux arguments on n'a pas besoin de construire chaque fois la surface $z = \varphi(x, y)$ comme sur la figure 12.4.1 ni de la couper par un plan parallèle à xOy. Il suffit de tracer dans le plan xOy une courbe d'équation $z = \varphi(x, y)$, de noter de quel côté de cette courbe on a $Z < z$ et de quel côté $Z > z$ et d'intégrer sur le domaine D où $Z < z$.

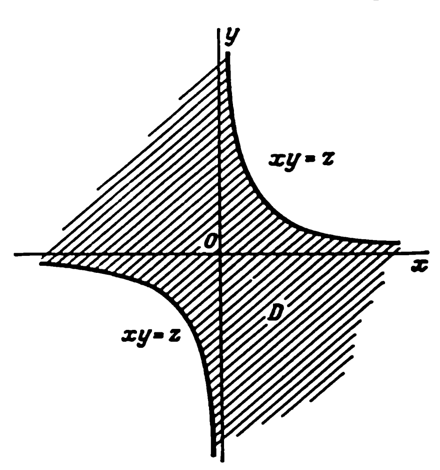


Fig. 12.4.2

E x e m p l e. Soit un système de variables (X, Y) de loi de répartition $f(x, y)$. Z est le produit des variables X, Y:

$$Z = XY.$$

Trouver la densité de probabilité de Z.

S o l u t i o n. Fixons-nous une certaine valeur z et construisons sur le plan xOy une courbe d'équation $xy = z$ (fig. 12.4.2). C'est une hyperbole ayant pour asymptotes les axes de coordonnées. Le domaine D sur la figure 12.4.2 est hachuré.

---

*) Sur notre figure la courbe C est fermée; en général, elle peut ne pas être fermée et se composer de plusieurs branches.

La fonction de répartition de Z s'écrit:

$$G(z)=\iint\limits_{(D)} f(x, y)\, dx\, dy=\int\limits_{-\infty}^{0}\int\limits_{z/x}^{\infty} f(x, y)\, dx\, dy+\int\limits_{0}^{\infty}\int\limits_{-\infty}^{z/x} f(x, y)\, dx\, dy .$$

En dérivant cette expression par rapport à z on a:

$$g(z)=-\int\limits_{-\infty}^{0}\frac{1}{x} f\left(x, \frac{z}{x}\right) dx+\int\limits_{0}^{\infty}\frac{1}{x} f\left(x, \frac{z}{x}\right) dx . \qquad (12.4.3)$$

12.5. Loi de répartition de la somme de deux variables aléatoires. Composition des lois de répartition

Nous allons utiliser la méthode générale exposée ci-dessus pour tésoudre un problème particulier important pour les applications pratiques. Il s'agit de la recherche de la loi de répartition de la *somme de deux variables aléatoires.*

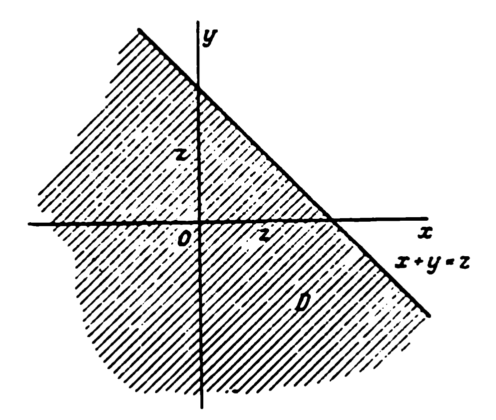


Fig. 12.5.1

Soit un système de deux variables aléatoires (X, Y) de densité de probabilité $f(x, y)$. Considérons la somme

$$Z = X + Y$$

pour laquelle nous allons trouver la loi de répartition. Pour cela traçons dans le plan xOy la droite d'équation $x + y = z$ (fig. 12.5.1). C'est une droite coupant sur les axes des segments de même longueur z et divisant le plan xOy en deux parties; à droite et au-dessus de cette droite $X + Y > z$; à gauche et au-dessous $X + Y < z$. Le domaine D est ici la partie inférieure gauche du plan xOy hachurée sur la figu-

re 12.5.1. En vertu de la formule (12.4.2) on a :

$$G(z)=\iint_{(D)} f(x,\ y)\,dx\,dy=\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,\ y)\,dx\,dy=$$

$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{z-x} f(x,\ y)\,dy\right\}dx.$$

En dérivant cette expression par rapport à la variable z entrant dans la limite supérieure de l'intégrale interne on a :

$$g(z)=\int_{-\infty}^{\infty} f(x,\ z-x)\,dx. \qquad (12.5.1)$$

C'est une forme générale pour la densité de probabilité d'une somme de deux variables aléatoires.

Vu que le problème est symétrique par rapport à X et Y on peut écrire cette formule sous la forme :

$$g(z)=\int_{-\infty}^{\infty} f(z-y,\ y)\,dy, \qquad (12.5.2)$$

équivalente à la première et pouvant être utilisée à sa place.

Pour les applications pratiques le cas des variables aléatoires $(X,\ Y)$ indépendantes est très important. On parle alors de *composition des lois de répartition* (ou *de convolution*).

Composer deux lois de répartition veut dire trouver la loi de répartition que suit la somme de deux variables aléatoires indépendantes en question.

Nous allons trouver la formule de composition de deux lois de répartition. Soient x et y deux variables aléatoires indépendantes de lois $f_1(x)$ et $f_2(y)$; trouver la composition de ces lois, c'est-à-dire trouver la densité de probabilité de la grandeur

$$Z=X+Y.$$

Les variables X et Y étant indépendantes, on a:

$$f(x,\ y)=f_1(x)\,f_2(y)$$

et les formules (12.5.1) et (12.5.2) s'écrivent alors :

$$g(z)=\int_{-\infty}^{\infty} f_1(x)f_2(z-x)\,dx, \qquad (12.5.3)$$

$$(z)=\int_{-\infty}^{\infty} f_1(z-y)\,f_2(y)\,dy. \qquad (12.5.4)$$

Pour désigner la composition des lois de répartition on utilise souvent la notation symbolique:

$$g = f_1 * f_2,$$

où $*$ est le symbole de composition (de convolution).

Les formules (12.5.3) et (12.5.4) ne sont commodes que si les lois $f_1(x)$ et $f_2(y)$ (ou tout au moins l'une d'elles) sont données par une formule unique dans toute la gamme des valeurs de l'argument (de $-\infty$ à ∞). Si au contraire les deux lois sont données dans les différents domaines par des équations différentes (par exemple, deux lois de densité uniforme) il est plus commode d'utiliser directement la méthode générale exposée au § 12.4, c'est-à-dire de calculer la fonction de répartition $G(z)$ de la grandeur $Z = X + Y$ et de prendre la dérivée de cette fonction.

Exemple 1. Réaliser la composition de la loi normale

$$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

et de la loi uniforme

$$f_2(y) = \frac{1}{\beta-\alpha} \quad \text{pour} \quad \alpha < y < \beta.$$

Solution. Nous allons appliquer la formule de composition des lois de répartition (12.5.4):

$$g(z) = \frac{1}{\beta-\alpha} \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-y-m)^2}{2\sigma^2}} dy = \frac{1}{\beta-\alpha} \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y-(z-m)]^2}{2\sigma^2}} dy. \tag{12.5.5}$$

La fonction sous l'intégrale dans l'expression (12.5.5) n'est rien d'autre qu'une loi normale dont le centre de dispersion est $z - m$ et l'écart quadratique moyen σ, l'intégrale dans l'expression (12.5.5) est la probabilité pour la variable aléatoire suivant cette loi de se trouver dans l'intervalle de α à β; par conséquent:

$$g(z) = \frac{1}{\beta-\alpha}\left[\Phi^*\left(\frac{\beta-(z-m)}{\sigma}\right) - \Phi^*\left(\frac{\alpha-(z-m)}{\sigma}\right)\right].$$

Sur la figure 12.5.2 on peut trouver les graphiques des lois $f_1(x)$, $f_2(y)$ et $g(z)$ pour $\alpha = -2$, $\beta = 2$, $m = 0$, $\sigma = 1$.

Exemple 2. Trouver le produit de composition des deux lois suivantes de densités uniformes définies sur le même intervalle (0, 1):

$$f_1(x) = 1 \text{ pour } 0 < x < 1,$$
$$f_2(y) = 1 \text{ pour } 0 < y < 1.$$

Solution. Les lois $f_1(x)$ et $f_2(y)$ n'étant données que sur certains intervalles des axes Ox et Oy, pour résoudre ce problème il vaut mieux d'utiliser la méthode générale exposée dans le § 12.4 et non pas les formules (12.5.3)

et (12.5.4) et trouver la fonction de répartition $G(z)$ de la grandeur $Z = X + Y$.

Considérons le point aléatoire (X, Y) dans le plan xOy. Le domaine où peut se trouver ce point est le carré R de côté égal à 1 (fig. 12.5.3). On a:

$$G(z) = \iint_{(D)} f(x, y)\, dx\, dy = \iint_{(D)} dx\, dy,$$

où le domaine D est la partie du carré R se trouvant à gauche et au-dessous de la droite $x + y = z$. Il est évident que

$$G(z) = S_D,$$

où S_D est l'aire du domaine D.

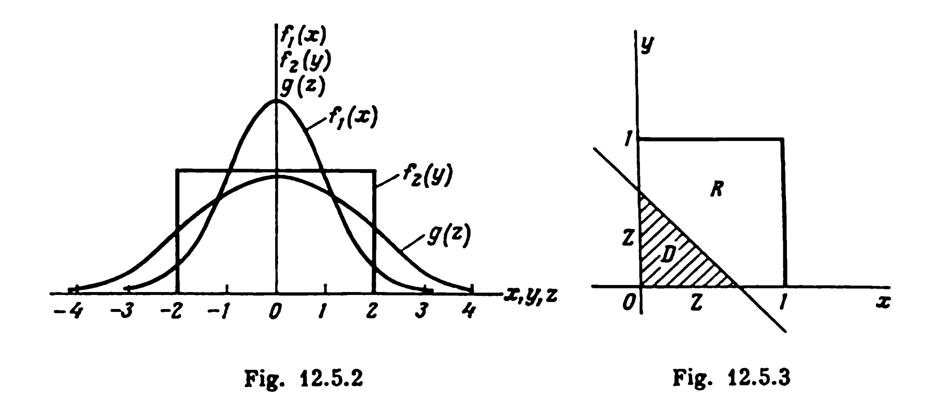


Fig. 12.5.2

Fig. 12.5.3

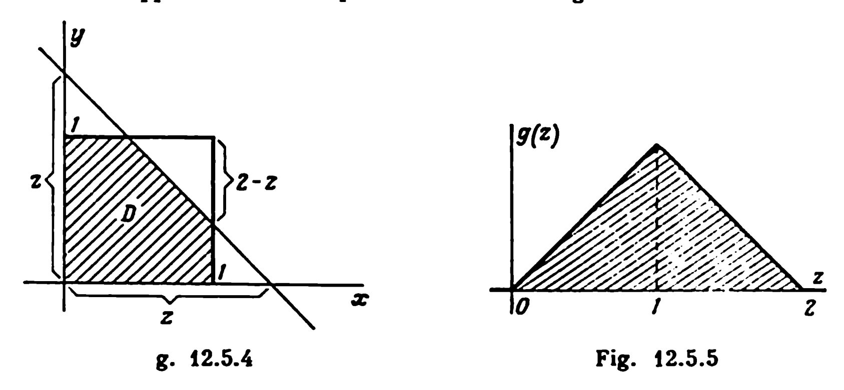
Ecrivons l'expression de l'aire du domaine D pour différentes valeurs de z en utilisant à cette fin les figures 12.5.3 et 12.5.4:

z	$G(z)$
$z < 0$	$G(z) = 0$
$0 < z < 1$	$G(z) = \frac{z^2}{2}$
$1 < z < 2$	$G(z) = 1 - \frac{(2-z)^2}{2}$
$z > 2$	$G(z) = 1$

En dérivant ces expressions on obtient:

z	$g(z)$
$z<0$	$g(z)=0$
$0<z<1$	$g(z)=z$
$1<z<2$	$g(z)=2-z$
$z>2$	$g(z)=0$

Le graphique de la loi de répartition $g(z)$ est donné sur la figure 12.5.5. Cette loi s'appelle loi de Simpson ou loi du triangle.



g. 12.5.4

Fig. 12.5.5

12.6. Composition des lois normales

Soient X et Y deux variables aléatoires indépendantes réparties suivant les lois normales:

$$f_1(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}},\qquad (12.6.1)$$

$$f_2(y)=\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}.\qquad (12.6.2)$$

Il y a lieu d'effectuer la composition de ces lois, c'est-à-dire de trouver la loi de répartition de la grandeur

$$Z=X+Y.$$

Appliquons la formule générale (12.5.3):

$$g(z)=\int_{-\infty}^{\infty} f_1(x)\,f_2(z-x)\,dx=\frac{1}{2\pi\sigma_x\sigma_y}\int_{-\infty}^{\infty} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}-\frac{(z-x-m_y)^2}{2\sigma_y^2}}\,dx. \tag{12.6.3}$$

En ouvrant les parenthèses dans l'exposant de la fonction sous l'intégrale et en simplifiant on obtient:

$$g(z)=\frac{1}{2\pi\sigma_x\sigma_y}\int_{-\infty}^{\infty} e^{-Ax^2+2Bx-C}\,dx,$$

où

$$A=\frac{1}{2}\,\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2\sigma_y^2};$$

$$B=\frac{m_x}{2\sigma_x^2}+\frac{z-m_y}{2\sigma_y^2};$$

$$C=\frac{m_x^2}{2\sigma_x^2}+\frac{(z-m_y)^2}{2\sigma_y^2}.$$

Substituons ces expressions dans la formule (9.1.3), il vient:

$$\int_{-\infty}^{\infty} e^{-Ax^2\pm 2Bx-C}\,dx=\sqrt{\frac{\pi}{A}}\,e^{-\frac{AC-B^2}{A}}, \tag{12.6.4}$$

et, après transformations:

$$g(z)=\frac{1}{\sqrt{\sigma_x^2+\sigma_y^2}}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{[z-(m_x+m_y)]^2}{2(\sigma_x^2+\sigma_y^2)}}, \tag{12.6.5}$$

ce n'est rien d'autre qu'une loi normale de centre de dispersion

$$m_z=m_x+m_y \tag{12.6.6}$$

et d'écart quadratique moyen

$$\sigma_z=\sqrt{\sigma_x^2+\sigma_y^2}. \tag{12.6.7}$$

Les raisonnements qualitatifs suivants permettent d'arriver à la même conclusion bien plus rapidement.

Sans ouvrir les parenthèses ni effectuer les transformations dans la fonction sous l'intégrale de (12.6.3) on voit immédiatement que l'exposant est un trinôme en x de la forme:

$$\varphi(x)=-Ax^2+2Bx-C,$$

où le coefficient A ne contient pas z, le coefficient B lui est proportionnel, et le coefficient C proportionnel à son carré. Utilisant la

formule (12.6.4) on voit ainsi que $g(z)$ est une exponentielle dont l'exposant est un trinôme en z ; or, la densité de probabilité de cette forme correspond à la loi normale. On arrive ainsi à une conclusion purement qualitative : la loi de répartition de la variable z doit être normale.

Pour trouver les paramètres de cette loi, à savoir m_z et σ_z, nous allons utiliser le théorème de composition des espérances mathématiques et le théorème de composition des variances. En vertu du premier théorème :

$$m_z = m_x + m_y, \tag{12.6.8}$$

et en vertu du second :

$$D_z = D_x + D_y,$$

ou

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2, \tag{12.6.9}$$

d'où la formule (12.6.7).

En passant des écarts quadratiques moyens aux écarts probables qui leur sont proportionnels on obtient :

$$E_z^2 + E_x^2 + E_y^2. \tag{12.6.10}$$

On arrive ainsi à la règle suivante : *la composition des lois normales est une loi normale, les espérances mathématiques et les variances (ou les carrés des écarts probables) étant la somme des grandeurs correspondantes.*

La règle de composition des lois normales peut être généralisée au cas d'un nombre arbitraire de variables aléatoires indépendantes.

Soient :

$$X_1, \ X_2, \ \ldots, \ X_n$$

n variables aléatoires indépendantes normales dont les moyennes sont :

$$m_{x_1}, \quad m_{x_2}, \ \ldots, \ m_{x_n}$$

et les écarts quadratiques moyens :

$$\sigma_{x_1}, \quad \sigma_{x_2}, \ \ldots, \ \sigma_{x_n} ;$$

la grandeur

$$Z = \sum_{i=1}^{n} X_i$$

suit également une loi normale de paramètres :

$$m_z = \sum_{i=1}^{n} m_{x_i}, \tag{12.6.11}$$

$$\sigma_z^2 = \sum_{i=1}^{n} \sigma_{x_i}^2. \tag{12.6.12}$$

Au lieu de la formule (12.6.12) on peut utiliser la formule équivalente :

$$E_z^2=\sum_{i=1}^{n} E_{x_i}^2. \tag{12.6.13}$$

Si le système de variables aléatoires (X, Y) est réparti suivant une loi normale mais X et Y sont liées, on peut comme auparavant montrer que, partant de la formule générale (12.5.1), la loi de répartition de la grandeur :

$$Z = X + Y$$

est également normale. Les moyennes s'ajoutent algébriquement, mais pour les écarts quadratiques moyens la règle de composition est plus compliquée ; on a :

$$\sigma_z^2=\sigma_x^2+\sigma_y^2+2r\sigma_x\sigma_y, \tag{12.6.14}$$

où r est le coefficient de corrélation des variables X et Y.

Lorsque l'on ajoute plusieurs variables aléatoires interdépendantes normales, la loi de répartition de la somme est également une loi normale de paramètres :

$$m_z=\sum_{i=1}^{n} m_{x_i}, \tag{12.6.15}$$

$$\sigma_z^2=\sum_{i=1}^{n} \sigma_{x_i}^2+2\sum_{i<j} r_{ij}\sigma_{x_i}\sigma_{x_j}, \tag{12.6.16}$$

ou pour les écarts probables

$$E_z^2=\sum_{i=1}^{n} E_{x_i}^2+2\sum_{i<j} r_{ij}E_{x_i}E_{x_j}, \tag{12.6.17}$$

où r_{ij} est le coefficient de corrélation des variables X_i, X_j, la sommation s'étendant à toutes les combinaisons deux à deux des variables $X_1, X_2, \ldots, X_n$.

Nous venons de voir une propriété très importante de la loi normale : la composition de lois normales est une loi normale. C'est la propriété dite *de stabilité*. Une loi de répartition est dite *stable* si lors de la composition de deux lois de ce type on obtient une loi de même type *). Ci-dessus nous avons montré que la loi normale est stable. Peu de lois de répartition sont douées de la propriété de stabilité. Dans le précédent paragraphe (exemple 2) nous avons vu par exemple que la loi uniforme n'est pas stable : par composition de deux lois uniformes données dans les intervalles de 0 à 1 on obtient la loi de Simpson.

---

*) On entend par lois du même type, des lois différant les unes des autres par les échelles et l'origine sur l'axe des abscisses.

La loi normale doit sa grande popularité à cette propriété de stabilité. Cependant, d'autres lois sont également douées de cette propriété. La loi normale a ceci de particulier que lors de la composition d'un nombre suffisamment grand de lois arbitraires, la « loi produit » est aussi voisine que l'on veut de la loi normale, quelles que soient les lois de répartition des composantes. On peut montrer ceci sur l'exemple de la composition de trois lois uniformes données sur les intervalles de 0 à 1. La loi de répartition $g(z)$ ainsi obtenue est montrée sur la figure 12.6.1. On voit que le graphique de la fonction $g(z)$ rappelle la loi normale.

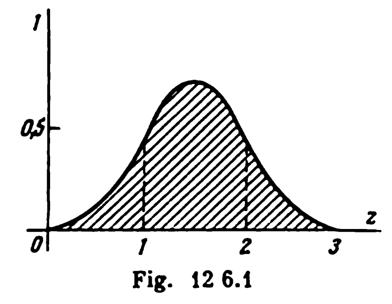


Fig. 12 6.1

12.7. Fonctions linéaires d'arguments normalement répartis

Soit un système de variables aléatoires $(X_1, X_2, \ldots, X_n)$ normales ; la grandeur aléatoire Y est une fonction linéaire de ces variables :

$$Y = \sum_{i=1}^{n} a_i X_i + b. \tag{12.7.1}$$

Il y a lieu de trouver la loi de répartition de Y.

Il est facile de voir que c'est une loi normale. En effet, la grandeur Y est une somme de fonctions linéaires dont chacune dépend d'un seul argument X normal, et comme nous l'avons montré ci-dessus, cette fonction linéaire suit également une loi normale. En ajoutant plusieurs grandeurs normalement réparties on obtient de nouveau une grandeur normale.

Il ne reste plus qu'à trouver les paramètres de la loi de Y, à savoir, la moyenne m_y et l'écart quadratique moyen σ_y. En appliquant les théorèmes de l'espérance mathématique et de la variance on obtient :

$$m_y = \sum_{i=1}^{n} a_i m_{x_i} + b, \tag{12.7.2}$$

$$\sigma_y^2 = \sum_{i=1}^{n} a_i^2 \sigma_{x_i}^2 + 2 \sum_{i<j} a_i a_j r_{ij} \sigma_{x_i} \sigma_{x_j}, \tag{12.7.3}$$

où r_{ij} est le coefficient de corrélation des grandeurs X_i, X_j.

Lorsque les variables $(X_1, X_2, \ldots, X_n)$ ne sont pas corrélées (donc, indépendantes pour la loi normale) la formule (12.7.3) devient :

$$\sigma_y^2 = \sum_{i=1}^{n} a_i^2 \sigma_{x_i}^2. \tag{12.7.4}$$

Dans les formules (12.7.3) et (12.7.4) les écarts quadratiques moyens peuvent être remplacés par les écarts probables qui leur sont proportionnels.

Dans les applications pratiques souvent on rencontre des cas où l'on ne connaît pas exactement les lois de répartition des variables aléatoires $X_1, X_2, \ldots, X_n$ figurant dans la formule (12.7.1), mais seulement leurs caractéristiques numériques, à savoir l'espérance mathématique et la variance. Si les variables $X_1, X_2, \ldots, X_n$ sont de plus indépendantes et leur nombre suffisamment important, on peut affirmer que quelles que soient les lois de répartition des X_i, la grandeur Y suit une loi voisine de la loi normale. En fait, pour obtenir la loi de répartition pouvant être approximativement considérée comme normale, il suffit, en général, de 5 à 10 composantes dans l'expression (12.7.1). Ceci n'est plus vrai lorsque la variance de l'une des composantes dans la formule (12.7.1) prédomine; on suppose que les variances des termes de la somme (12.7.1) sont d'un même ordre de grandeur. Ces conditions étant observées, on peut adopter approximativement pour la grandeur Y une loi normale dont les paramètres sont donnés par les formules (12.7.2) et (12.7.4).

Il est évident que tous les raisonnements ci-dessus concernant la loi de répartition d'une fonction linéaire sont vrais (bien sûr, approximativement) lorsque la fonction n'est pas tout à fait linéaire mais peut être linéarisée.

Exemple. Pour faire son point un avion mesure la distance le séparant de deux repères se trouvant environ à 50 km à l'ouest et à 100 km au sud-ouest. Les erreurs de mesure sont indépendantes et caractérisées par un écart quadratique moyen égal à 100 m; les erreurs systématiques sont exclues. Trouver les paramètres de la loi de répartition de l'erreur de position de l'avion.

Solution. Prenons l'axe des abscisses dirigé vers l'est et celui des ordonnées vers le nord. Désignons par (a_1, b_1), (a_2, b_2) les coordonnées des repères; la position de l'avion est représentée par le point (x, y). Les distances vraies aux repères sont respectivement $\sqrt{(x-a_1)^2+(y-b_1)^2}$, $\sqrt{(x-a_2)^2+(y-b_2)^2}$; les distances mesurées sont égales aux distances vraies plus l'erreur:

$$\begin{aligned} R_1 &= \sqrt{(x-a_1)^2+(y-b_1)^2}+\Delta R_1, \\ R_2 &= \sqrt{(x-a_2)^2+(y-b_2)^2}+\Delta R_2, \end{aligned} \tag{12.7.5}$$

ΔR_1 et ΔR_2 sont ici indépendants.

Les équations donnant X, Y de l'avion en fonction des distances mesurées sont:

$$\begin{cases} (x-a_1)^2+(y-b_1)^2=R_1^2, \\ (x-a_2)^2+(y-b_2)^2=R_2^2, \end{cases}$$

ou compte tenu de (12.7.5),

$$\begin{cases} (x-a_1)^2+(y-b_1)^2=(x-a_1)^2+(y-b_1)^2+ \\ \qquad +2\sqrt{(x-a_1)^2+(y-b_1)^2}\cdot\Delta R_1+(\Delta R_1)^2, \\ (x-a_2)^2+(y-b_2)^2=(x-a_2)^2+(y-b_2)^2+ \\ \qquad +2\sqrt{(x-a_2)^2+(y-b_2)^2}\cdot\Delta R_2+(\Delta R_2)^2. \end{cases} \tag{12.7.6}$$

Ecrivons les coordonnées X, Y données par le système (12.7.6) comme suit: $X = x + \Delta X$, $Y = y + \Delta Y$, ΔX, ΔY étant les erreurs, et linéarisons le système. Après simplifications nous avons:

$$\begin{cases} 2(x-a_1)\cdot\Delta X + 2(y-b_1)\cdot\Delta Y \approx 2\sqrt{(x-a_1)^2+(y-b_1)^2}\cdot\Delta R_1, \\ 2(x-a_2)\cdot\Delta X + 2(y-b_2)\,\Delta Y \approx 2\sqrt{(x-a_2)^2+(y-b_2)^2}\cdot\Delta R_2. \end{cases} \quad (12.7.7)$$

Comme nous connaissons approximativement les directions des repères et leurs distances, nous pouvons écrire:

$$x - a_1 \approx 50 \text{ (km)}, \quad y - b_1 \approx 0,$$
$$x - a_2 \approx y - b_2 \approx 50\sqrt{2} \text{ (km)}.$$

Le système (12.7.7) s'écrit alors comme suit:

$$\begin{cases} 100\Delta X \approx 100\Delta R_1, \\ 100\sqrt{2}\,\Delta X + 100\sqrt{2}\,\Delta Y \approx 200\Delta R_2. \end{cases}$$

d'où:

$$\Delta X \approx \Delta R_1,$$
$$\Delta Y \approx \sqrt{2}\,\Delta R_2 - \Delta R_1.$$

Les espérances mathématiques des erreurs sont nulles (erreurs systématiques absentes). Les éléments de la matrice de corrélation sont

$$\sigma_{\Delta x}^2 = \sigma_{r_1}^2 = 10\,000, \quad \sigma_{\Delta x} = 100 \text{ (m)};$$
$$\sigma_{\Delta y}^2 = 2\sigma_{r_2}^2 + \sigma_{r_1}^2 - 30\,000, \quad \sigma_{\Delta y} \approx 173 \text{ (m)};$$
$$K_{\Delta x \Delta y} = -\sigma_{r_1}^2 = -10\,000.$$

Le coefficient de corrélation est égal à:

$$r_{\Delta x \Delta y} = -\frac{10\,000}{100\cdot 100\sqrt{3}} = -\frac{1}{\sqrt{3}} \approx -0\ 577.$$

12.8. Composition des lois normales dans le plan

Soient deux vecteurs aléatoires indépendants donnés dans le système xOy: $\vec{V}_1$ de composantes (X_1, Y_1) et $\vec{V}_2$ de composantes (X_2, Y_2) Supposons que chacun d'eux soit normalement réparti, les paramètres du premier vecteur étant m_{x_1}, m_{y_1}, σ_{x_1}, σ_{y_1}, $r_{x_1y_1}$ et ceux du second m_{x_2}, m_{y_2}, σ_{x_2}, σ_{y_2}, $r_{x_2y_2}$. Trouver la loi de répartition du vecteur $\vec{V} = \vec{V}_1 + \vec{V}_2$ (fig. 12.8.1) de composantes

$$X = X_1 + X_2;$$
$$Y = Y_1 + Y_2.$$

Il est facile de démontrer qualitativement (comme nous l'avons fait pour la composition de deux lois normales au § 12.6) que le vecteur $\vec{V}$ est également normalement réparti. Adoptons cette assertion sans démonstration.

Nous allons maintenant trouver les paramètres de la loi de répartition du vecteur $\vec{V}$. En vertu du théorème de la somme des espérances mathématiques on a:

$$\left.\begin{aligned} m_x &= m_{x_1} + m_{x_2}, \\ m_y &= m_{y_1} + m_{y_2}. \end{aligned}\right\} \tag{12.8.1}$$

Puis, en vertu du théorème de la somme des variances:

$$\left.\begin{aligned} \sigma_x^2 &= \sigma_{x_1}^2 + \sigma_{x_2}^2, \\ \sigma_y^2 &= \sigma_{y_1}^2 + \sigma_{y_2}^2. \end{aligned}\right\} \tag{12.8.2}$$

Et enfin, en vertu du théorème de la somme des covariances:

$$K_{xy} = K_{x_1y_1} + K_{x_2y_2},$$

ou, en passant aux coefficients de corrélation:

$$r_{xy}\sigma_x\sigma_y = r_{x_1y_1}\sigma_{x_1}\sigma_{y_1} + r_{x_2y_2}\sigma_{x_2}\sigma_{y_2},$$

d'où

$$r_{xy} = \frac{r_{x_1y_1}\sigma_{x_1}\sigma_{y_1} + r_{x_2y_2}\sigma_{x_2}\sigma_{y_2}}{\sqrt{(\sigma_{x_1}^2 + \sigma_{x_2}^2)(\sigma_{y_1}^2 + \sigma_{y_2}^2)}}. \tag{12.8.3}$$

Ainsi les formules (12.8.1), (12.8.2) et (12.8.3) permettent de résoudre le problème de composition des lois normales.

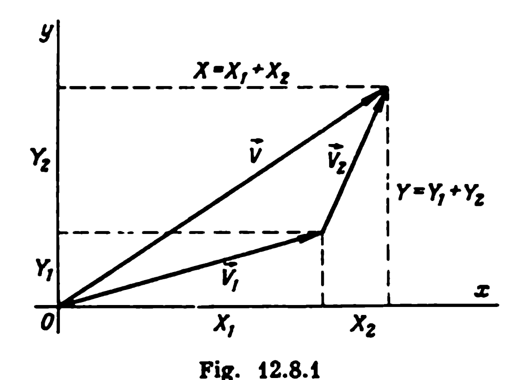


Fig. 12.8.1

Ces formules ont été démontrées pour le cas où les deux lois normales initiales (pour les vecteurs $\vec{V}_1$ et $\vec{V}_2$) sont données dans le même système de coordonnées xOy. Parfois il y a lieu de composer deux lois normales dans le plan, chacune étant donnée dans son propre système de coordonnées, à savoir dans ses axes principaux de l'ellipse de dispersion. Nous allons donner une méthode de composition des lois normales pour ce cas.

Supposons que, dans le plan xOy (fig. 12.8.2) on ait deux vecteurs aléatoires non corrélés normalement répartis $\vec{V}_1$ et $\vec{V}_2$. Chacun de ces vecteurs est caractérisé par son ellipse unitaire de dispersion; le vecteur $\vec{V}_1$ par l'ellipse de centre au point m_{x_1}, m_{y_1} et de demi-axes σ_{ξ_1}, σ_{η_1} dont le premier forme avec l'axe Ox l'anglé α_1 ; les caractéristiques analogues du vecteur $\vec{V}_2$ seront $m_{x_2}, m_{y_2}, \sigma_{\xi_2}, \sigma_{\eta_2}, \alpha_2$. Trouver

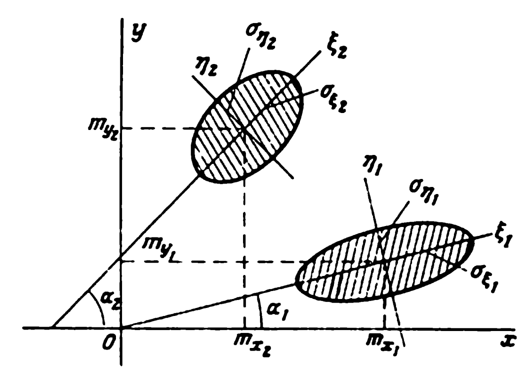


Fig. 12.8.2

les paramètres de l'ellipse unitaire de dispersion caractérisant le vecteur $\vec{V} = \vec{V}_1 + \vec{V}_2$. Nous désignerons ces paramètres

$$m_x,\ m_y,\ \sigma_x,\ \sigma_y,\ \alpha.$$

La position du centre de dispersion ne dépendant pas du choix du système de coordonnées, on aura donc comme auparavant les relations:

$$m_x = m_{x_1} + m_{x_2},$$

$$m_y = m_{y_1} + m_{y_2}.$$

Pour trouver les éléments de la matrice de corrélation du vecteur $\vec{V}$, nous allons projeter les points aléatoires correspondant aux vecteurs $\vec{V}_1$ et $\vec{V}_2$ sur les axes Ox et Oy. En vertu des formules (10.3.3):

$$\left.\begin{aligned}
\sigma_{x_1}^2 &= \sigma_{\xi_1}^2 \cos^2\alpha_1 + \sigma_{\eta_1}^2 \sin^2\alpha_1,\\
\sigma_{y_1}^2 &= \sigma_{\xi_1}^2 \sin^2\alpha_1 + \sigma_{\eta_1}^2 \cos^2\alpha_1,\\
\sigma_{x_2}^2 &= \sigma_{\xi_2}^2 \cos^2\alpha_2 + \sigma_{\eta_2}^2 \sin^2\alpha_2,\\
\sigma_{y_2}^2 &= \sigma_{\xi_2}^2 \sin_2\alpha_2 + \sigma_{\eta_2}^2 \cos^2\alpha_2.
\end{aligned}\right\} \qquad (12.8.4)$$

Les coefficients de corrélation des composantes des vecteurs $\vec{V}_1$ et $\vec{V}_2$ dans le système de coordonnées xOy sont donnés par la relation

(9.2.2):

$$\left.\begin{aligned} r_{x_1y_1} &= \frac{\operatorname{tg} 2\alpha_1 (\sigma_{x_1}^2 - \sigma_{y_1}^2)}{2\sigma_{x_1}\sigma_{y_1}}, \\ r_{x_2y_2} &= \frac{\operatorname{tg} 2\alpha_2 (\sigma_{x_2}^2 - \sigma_{y_2}^2)}{2\sigma_{x_2}\sigma_{y_2}}. \end{aligned}\right\} \quad (12.8.5)$$

La composition des lois normales dans le plan revient ainsi au problème précédent. Connaissant σ_x, σ_y, r_{xy} on peut à l'aide de la formule (9.2.2) trouver les angles formés par les axes de l'ellipse somme avec l'axe des abscisses:

$$\operatorname{tg} 2\alpha = \frac{2r_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \text{ *)}, \quad (12.8.6)$$

et à l'aide des formules (9.2.4) les écarts quadratiques moyens:

$$\left.\begin{aligned} \sigma_\xi &= \sqrt{\sigma_x^2 \cos^2\alpha + r_{xy}\sigma_x\sigma_y \sin 2\alpha + \sigma_y^2 \sin^2\alpha}, \\ \sigma_\eta &= \sqrt{\sigma_x^2 \sin^2\alpha - r_{xy}\sigma_x\sigma_y \sin 2\alpha + \sigma_y^2 \cos^2\alpha}. \end{aligned}\right\} \quad (12.8.7)$$

Les dernières relations sont vraies non seulement pour les écarts quadratiques moyens mais également pour les écarts probables qui leur sont proportionnels:

$$\begin{aligned} E_\xi &= \sqrt{E_x^2 \cos^2\alpha + r_{xy}E_xE_y \sin 2\alpha + E_y^2 \sin^2\alpha}, \\ E_\eta &= \sqrt{E_x^2 \sin^2\alpha - r_{xy}E_xE_y \sin 2\alpha + E_y^2 \cos^2\alpha}. \end{aligned} \quad (12.8.8)$$

Nous allons passer maintenant à la composition d'un nombre quelconque de lois normales dans le plan.

Le cas le plus simple de composition d'un nombre quelconque de lois normales est celui où les axes principaux de dispersion pour toutes les lois entrant dans la composition sont parallèles. En choisissant les axes de coordonnées parallèlement à ces axes principaux de dispersion, nous aurons alors des systèmes de variables aléatoires indépendantes, la composition des lois normales sera donnée par les formules simples suivantes:

$$\left.\begin{aligned} m_x &= \sum_{i=1}^{n} m_{x_i}; \qquad & \sigma_x^2 &= \sum_{i=1}^{n} \sigma_{x_i}^2; \\ m_y &= \sum_{i=1}^{n} m_{y_i}; \qquad & \sigma_y^2 &= \sum_{i=1}^{n} \sigma_{y_i}^2, \end{aligned}\right\} \quad (12.8.9)$$

*) La tangente ayant pour période π, les valeurs de α données par la formule (12.8.6) peuvent différer de $\frac{\pi}{2}$, ce qui correspond à deux axes principaux de l'ellipse.

où σ_x, σ_{x_i}, σ_y, σ_{y_i} sont les écarts quadratiques moyens principaux des lois correspondantes.

Lorsque les directions des axes principaux diffèrent, on peut trouver la composition de plusieurs lois normales par la méthode utilisée pour le cas de deux lois, c'est-à-dire en projetant les vecteurs aléatoires à additionner sur les axes d'un même système de coordonnées.

Dans les applications on rencontre souvent des lois dites dégénérées, c'est-à-dire des lois caractérisées par une ellipse de dispersion n'ayant qu'un seul demi-axe, l'autre étant nul. Ces lois dégénérées donnent la dispersion dans une seule direction. Lors de la composition de ces lois il y a lieu de procéder comme dans le cas des lois ordinaires et supposer certains paramètres (écarts moyens quadratiques ou probables) nuls

Exemple 1. Un parachutiste saute d'un aérostat et essaie d atterrir en un certain point. L'erreur du point d'atterrissage se compose

1) de l'erreur liée à ce que la direction et la vitesse du vent sont différentes aux différentes altitudes;

2) des erreurs dues à d'autres causes.

Ces deux groupes d'erreurs sont indépendants. Les axes principaux de la dispersion, due au fait que l'on a mal tenu compte du vent, sont dirigés dans le sens du vent et perpendiculairement, les écarts quadratiques moyens principaux sont respectivement 20 et 10 m; les erreurs dues à d'autres causes donnent une ellipse unitaire qui est un cercle de rayon égal à 15 m. Les erreurs systématiques sont exclues. Trouver les paramètres de la loi normale que suit l'erreur totale du point d'atterrissage.

Solution. Nous allons diriger l'axe Ox suivant la direction dominante du vent et l'axe Oy, perpendiculairement. En appliquant la règle de composition des lois normales à composantes indépendantes on a [formules (12.8.9)]:

$$m_x = m_y = 0$$

$$\sigma_x^2 = 20^2 + 15^2 = 625, \qquad \sigma_x = 25\text{m}$$

$$\sigma_y^2 = 10^2 + 15^2 = 325, \qquad \sigma_y \approx 18\text{m}$$

Exemple 2. Dans les conditions de l'exemple 1 du § 12.7 un avion jette sur la terre une caisse avec des appareils. La dispersion du point d'atterrissage de la caisse due à des causes techniques est circulaire, l'écart quadratique moyen étant égal à 70 m et l'erreur systématique étant 30 m déplacée vers le nord (direction du vol de l'avion). Trouver les paramètres de la loi de répartition de l'erreur $(\Delta X, \Delta Y)$ du point d'atterrissage de la caisse due à des causes techniques et à l'imprécision de la détermination des coordonnées de l'avion; rapporter cette répartition aux axes principaux.

Solution. Désignons par ΔX_1, ΔY_1 l'erreur due à l'imprécision des coordonnées; par ΔX_2, ΔY_2 l'erreur due à des causes techniques. Les paramètres de ces composantes sont:

$$m_{\Delta x_1} = m_{\Delta y_1} = m_{\Delta x_2} = 0, \qquad m_{\Delta y_2} = 30;$$

$$\sigma_{\Delta x_1}^2 = 10\,000, \quad \sigma_{\Delta y_1}^2 = 30\,000, \qquad K_{\Delta x_1 \Delta y_1} = -10\,000;$$

$$\sigma_{\Delta x_2}^2 = \sigma_{\Delta y_2}^2 = 4\,900, \qquad K_{\Delta x_2 \Delta y_2} = 0.$$

En appliquant les théorèmes de la somme des espérances mathématiques, des variances et des covariances on obtient:

$$m_{\Delta x}=0, \qquad m_{\Delta y}=30;$$

$$\sigma^2_{\Delta x}=14\,900, \qquad \sigma_{\Delta x}\approx 122 \text{ (m)};$$

$$\sigma^2_{\Delta y}=34\,900, \qquad \sigma_{\Delta y}\approx 187 \text{ (m)}.$$

$$K\Delta x\Delta y=-10\,000.$$

Le coefficient de corrélation est:

$$r_{\Delta x\Delta y}=\frac{-10\,000}{122\cdot 187}\approx -0{,}44.$$

Calculons l'angle α formé par l'axe des abscisses et le premier axe principal de dispersion:

$$\operatorname{ctg} 2\alpha=\frac{2K_{\Delta x\Delta y}}{\sigma^2_{\Delta x}-\sigma^2_{\Delta y}}=\frac{-20\,000}{-20\,000}=1,$$

$$2\alpha = 45^\circ, \quad \alpha = 22^\circ 30'.$$

En vertu des formules (12.8.7) on a:

$$\sigma_\xi=\sqrt{\sigma^2_{\Delta x}\cos^2\alpha+K_{\Delta x\Delta y}\sin 2\alpha+\sigma^2_{\Delta y}\sin^2\alpha}\approx 104 \text{ (m)};$$

$$\sigma_\eta=\sqrt{\sigma^2_{\Delta x}\sin^2\alpha-K_{\Delta x\Delta y}\sin 2\alpha+\sigma^2_{\Delta y}\cos^2\alpha}\approx 198 \text{ (m)}.$$

Ainsi, le petit axe de l'ellipse de dispersion fait 22°30′ avec la direction est-ouest et le grand axe, 112°30′.

CHAPITRE 13

THÉORÈMES LIMITES DE LA THÉORIE DES PROBABILITÉS

13.1. Loi des grands nombres et théorème central limite

Au début de ce cours nous avons déjà noté que les lois mathématiques de la théorie des probabilités ont été obtenues en se basant sur les lois statistiques réelles propres aux phénomènes aléatoires d'attente. Ces lois sont liées au fait que dans les phénomènes d'attente on a un grand nombre d'expériences semblables ou la somme d'un grand nombre d'actions aléatoires, dont l'ensemble donne naissance à une grandeur aléatoire régie par une certaine loi. Dans l'antiquité on connaissait déjà la propriété de stabilité des phénomènes aléatoires ayant lieu en masse. Quelle que soit l'origine du phénomène, les particularités de chaque action aléatoire n'influent presque pas sur le résultat moyen de ces effets; des écarts aléatoires par rapport à la moyenne, inévitables pour chacun des effets, se trouvent compensés, nivelés, lissés dans leur masse. C'est cette stabilité de la moyenne qui représente le sens physique de la « loi des grands nombres » dans le sens général: lorsque le nombre des effets est très important leur résultat moyen cesse pratiquement d'être aléatoire et peut être prédit avec une très grande précision.

Dans le sens strict, en théorie des probabilités on entend par « loi des grands nombres » une série de théorèmes mathématiques, établissant chacun le fait que, pour telles ou telles conditions, les caractéristiques moyennes d'un grand nombre d'expériences tendent vers certaines constantes.

Dans le § 2.3 nous avons déjà formulé le théorème le plus simple de ce genre, à savoir le théorème de Bernoulli, en vertu duquel, la fréquence d'un événement, lors d'un grand nombre d'expériences, tend (ou, plus exactement, converge en probabilité) vers la probabilité de cet événement. Dans ce chapitre nous allons étudier des formules plus générales de la loi des grands nombres. Toutes ces formules établissent le fait et les conditions de convergence en probabilité de telles ou telles grandeurs aléatoires vers des grandeurs constantes, non aléatoires.

La loi des grands nombres joue un rôle important dans les applications pratiques de la théorie des probabilités. Le fait que les gran-

deurs aléatoires se comportent dans certaines conditions pratiquement comme des grandeurs constantes permet de les utiliser, de prédire les résultats des phénomènes aléatoires ayant lieu en masse avec certitude.

Les possibilités de ces pronostics concernant les phénomènes aléatoires de masse se trouvent élargies grâce aux théorèmes limites relatifs non pas aux valeurs limites des grandeurs aléatoires, mais aux lois de répartition limites. Il s'agit d'un groupe de théorèmes généralement appelés « théorème central limite ». Nous avons déjà noté que, sous certaines réserves, la loi de la somme d'un nombre suffisamment grand de variables aléatoires tend indéfiniment vers la loi normale. Ces conditions qui, du point de vue mathématique, peuvent se formuler différemment, sous une forme plus ou moins générale, signifient en fait que l'influence des composantes sur la somme est infiniment petite, c'est-à-dire que la somme ne contient pas de termes dont l'influence sur la dispersion de la somme est prédominante. Les différentes formes du théorème central limite diffèrent entre elles par les conditions pour lesquelles ces propriétés limites ont lieu.

Les différentes formes de la loi des grands nombres et celles du théorème central limite forment l'ensemble des *théorèmes limites* de la théorie des probabilités. Les théorèmes limites permettent non seulement de faire des pronostics scientifiques dans le domaine des phénomènes aléatoires, mais également d'estimer la précision de ces pronostics.

Dans ce chapitre nous n'envisagerons que les formes les plus simples des théorèmes limites. Nous étudierons d'abord les théorèmes du groupe de la « loi des grands nombres », puis les théorèmes du groupe du « théorème central limite ».

13.2. Inégalité de Tchébychev

Nous allons démontrer une inégalité très générale connue sous le nom d'*inégalité de Bienaymé-Tchébychev*, qui servira de lemme pour la démonstration des théorèmes du groupe de la « loi des grands nombres ».

Soit X une variable aléatoire d'espérance mathématique m_x et de variance D_x. Selon l'inégalité de Tchébychev pour tout nombre réel positif α, la probabilité pour que X s'écarte de son espérance mathématique d'une grandeur non inférieure à α, a pour limite supérieure $\frac{D_x}{\alpha^2}$:

$$P(|X-m_x|\geqslant\alpha)\leqslant\frac{D_x}{\alpha^2}. \qquad (13.2.1)$$

D é m o n s t r a t i o n. Soit X une variable discrète, répartie comme suit :

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Nous allons représenter les valeurs possibles de X et son espérance mathématique m_x par des points sur l'axe numérique Ox (fig. 13.2.1).

Fig. 13.2.1

Soit une certaine valeur $\alpha > 0$ et calculons la probabilité pour que X s'écarte de son espérance mathématique d'une grandeur non inférieure à α :

$$P(|X - m_x| \geqslant \alpha). \tag{13.2.2}$$

Pour cela nous portons à droite et à gauche du point m_x un segment de longueur α ; on obtient ainsi l'intervalle AB. La probabilité (13.2.2) n'est rien d'autre que la probabilité pour le point aléatoire X de se trouver à l'extérieur de l'intervalle AB :

$$P(|X - m_x| \geqslant \alpha) = P(X \not\subset AB)\ ^{*)}.$$

Cette probabilité est la somme des probabilités de tous les x_i se trouvant à l'extérieur de l'intervalle AB :

$$P(|X - m_x| \geqslant \alpha) = \sum_{|x_i - m_x| \geqslant \alpha} p_i, \tag{13.2.3}$$

où la notation $|x_i - m_x| \geqslant \alpha$ signifie que la sommation s'étend à tous les i pour lesquels les points x_i se trouvent à l'extérieur de l'intervalle AB.

D'autre part, écrivons l'expression de la variance de X. Par définition

$$D_x = M[(X - m_x)^2] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i = \sum_{i=1}^{n} |x_i - m_x|^2 p_i. \tag{13.2.4}$$

Tous les termes de la somme (13.2.4) étant non négatifs, cette dernière ne peut que diminuer lorsqu'on la calcule non pas pour toutes les valeurs de x_i, mais seulement pour certaines d'entre elles, en par-

*) Les extrémités de l'intervalle AB ne lui appartiennent pas.

ticulier, pour celles se trouvant à l'extérieur de l'intervalle AB:

$$D_x \geqslant \sum_{|x_i - m_x| \geqslant \alpha} |x_i - m_x|^2 p_i. \qquad (13.2.5)$$

Remplaçons dans la somme l'expression $|x_i - m_x|$ par α. Comme pour tous les termes de la somme on a $|x_i - m_x| \geqslant \alpha$, la substitution mentionnée ci-dessus ne peut que diminuer la somme; par conséquent:

$$D_x \geqslant \sum_{|x_i - m_x| \geqslant \alpha} \alpha^2 p_i = \alpha^2 \sum_{|x_i - m_x| \geqslant \alpha} p_i. \qquad (13.2.6)$$

Mais, en vertu de la formule (13.2.3), la dernière somme de (13.2.6) n'est rien d'autre que la probabilité pour le point aléatoire de se trouver à l'extérieur de l'intervalle AB; donc on a:

$$D_x \geqslant \alpha^2 P(|X - m_x| \geqslant \alpha),$$

ce qui démontre immédiatement l'inégalité en question.

2. Dans le cas où X est continue, la démonstration est analogue à condition de remplacer les p_i par des éléments de probabilité, et les sommes finies par des intégrales. En effet:

$$P(|X - m_x| > \alpha) = \int_{|x - m_x| > \alpha} f(x)\, dx\ ^{*)}, \qquad (13.2.7)$$

où $f(x)$ est la densité de probabilité de X. Ensuite

$$D_x = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)\, dx =$$

$$= \int_{-\infty}^{\infty} |x - m_x|^2 f(x)\, dx \geqslant \int_{|x - m_x| > \alpha} |x - m_x|^2 f(x)\, dx,$$

où $|x - m_x| > \alpha$ au-dessous de l'intégrale signifie que l'intégration s'étend à toute la partie extérieure de l'intervalle AB.

En remplaçant $|x - m_x|$ sous le signe de l'intégrale par α on obtient:

$$D_x \geqslant \alpha^2 \int_{|x - m_x| > \alpha} f(x)\, dx = \alpha^2 P(|X - m_x| > \alpha),$$

d'où on déduit l'inégalité de Tchébychev pour les variables continues.

E x e m p l e. Soit X une variable aléatoire d'espérance mathématique m_x et de variance σ_x^2. Trouver l'estimation majorante de la probabilité pour X de s'écarter de son espérance mathématique d'une grandeur au moins égale à $3\sigma_x$.

---

*) Le signe $\geqslant$ est remplacé par le signe $>$ car pour une variable continue la probabilité d'une égalité exacte est nulle.

S o l u t i o n. En posant $\alpha = 3\sigma_x$ dans l'inégalité de Tchébychev on a :

$$P(|X - m_x| \geqslant 3\sigma_x) \leqslant \frac{D_x}{9\sigma_x^2} = \frac{1}{9},$$

c'est-à-dire la probabilité pour qu'une variable aléatoire s'écarte de son espérance mathématique d'une grandeur dépassant trois fois l'écart quadratique moyen, ne peut être supérieure à 1/9.

N o t a. L'inégalité de Tchébychev ne donne que la limite supérieure de la probabilité de l'écart en question. Aucune loi de répartition ne peut donner une probabilité supérieure à cette valeur limite. En pratique dans la majorité des cas la probabilité pour que la variable X sorte des limites de l'intervalle $m_x \pm 3\sigma_x$ est nettement inférieure à 1/9. Par exemple, pour la loi normale cette probabilité est voisine de 0,003. En pratique, le plus souvent on a affaire à des variables aléatoires dont les valeurs sortent très rarement au-delà de $m_x \pm 3\sigma_x$. Si l'on ne connaît pas la loi de répartition d'une grandeur aléatoire, mais seulement m_x et σ_x, on considère que l'intervalle $m_x \pm 3\sigma_x$ est le domaine des valeurs possibles de la grandeur aléatoire (règle dite « des trois sigma »).

13.3. Loi des grands nombres (théorème de Tchébychev)

Dans ce paragraphe nous allons démontrer un théorème des plus simples mais à la fois des plus importants de la loi des grands nombres, à savoir le théorème de Tchébychev. Ce théorème établit une relation entre la moyenne arithmétique des valeurs observées d'une variable aléatoire et son espérance mathématique.

Préalablement nous allons résoudre le problème auxiliaire suivant.

Soit X une variable aléatoire d'espérance mathématique m_x et de variance D_x. Les n expériences indépendantes fournissent les valeurs observées de X dont on calcule la moyenne arithmétique. Il y a lieu de trouver les caractéristiques numériques de cette moyenne arithmétique, donc, son espérance mathématique et sa variance, et d'estimer leur variation avec augmentation de n.

Introduisons les désignations suivantes :

X_1, valeur de X dans la première expérience ;

X_2, valeur de X dans la seconde expérience, etc.

Il est évident que l'ensemble des valeurs $X_1, X_2, \ldots, X_n$ représente n variables aléatoires indépendantes réparties chacune suivant la même loi que la grandeur X. Soit la moyenne arithmétique de ces variables :

$$Y = \frac{\sum_{i=1}^{n} X_i}{n}$$

La variable aléatoire Y est une fonction linéaire des variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$. Calculons l'espérance mathématique et la variance de Y. En vertu des règles formulées dans le § 10 pour les caractéristiques numériques des fonctions linéaires on a :

$$m_y = M[Y] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n} n m_x = m_x;$$

$$D_y = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{D_x}{n}.$$

Ainsi, l'espérance mathématique de Y ne dépend pas du nombre d'expériences n et est égale à l'espérance mathématique de la variable observée X ; pour ce qui est de la variance de Y, elle décroît indéfiniment au fur et à mesure de l'augmentation du nombre d'expériences et peut être aussi petite que l'on veut lorsque n est suffisamment grand. On voit ainsi que la moyenne arithmétique est une variable aléatoire de variance aussi petite que l'on veut et pour un nombre d'expériences suffisamment grand elle se conduit *à peu près comme une grandeur non aléatoire.*

Le théorème de Tchébychev établit sous une forme qualitative exacte cette stabilité de la moyenne arithmétique. Ce théorème se formule comme suit :

Pour un nombre d'expériences indépendantes suffisamment grand, la moyenne arithmétique des valeurs observées d'une variable aléatoire converge en probabilité vers son espérance mathématique.

Nous allons écrire le théorème de Tchébychev sous la forme d'une formule. Rappelons-nous le sens de l'expression « convergence en probabilité ». On dit qu'une variable aléatoire X_n converge en probabilité vers a si lorsque n augmente, la probabilité pour que X_n et a soient aussi voisins que l'on veut tend indéfiniment vers l'unité, ce qui signifie que pour n suffisamment grand on a :

$$P(|X_n - a| < \varepsilon) > 1 - \delta,$$

où ε, δ sont des nombres infiniment petits arbitraires.

Ecrivons le théorème de Tchébychev sous une forme analogue. Selon ce théorème lorsque n augmente la moyenne arithmétique $\frac{\sum_{i=1}^{n} X_i}{n}$ converge en probabilité vers m_x, c'est-à-dire :

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| < \varepsilon\right) > 1 - \delta. \qquad (13.3.1)$$

Nous allons démontrer cette inégalité.

D é m o n s t r a t i o n. Nous avons démontré plus haut que la grandeur

$$Y = \frac{\sum_{i=1}^{n} X_i}{n}$$

a pour caractéristiques numériques

$$m_y = m_x; \quad D_y = \frac{D_x}{n}.$$

Appliquons à la grandeur Y l'inégalité de Tchébychev en posant $\alpha = \varepsilon$:

$$P(|Y - m_y| \geqslant \varepsilon) \leqslant \frac{D_y}{\varepsilon^2} = \frac{D_x}{n\varepsilon^2}.$$

Quelque petit que soit ε, on peut toujours trouver un n suffisamment grand pour que l'inégalité

$$\frac{D_x}{n\varepsilon^2} < \delta$$

soit remplie, δ étant un nombre aussi petit que l'on veut.

On a alors:

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| \geqslant \varepsilon\right) < \delta,$$

d'où, en passant à l'événement contraire, on a :

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| < \varepsilon\right) > 1 - \delta.$$

ce qu'il fallait démontrer.

13.4. Théorème de Tchébychev généralisé. Théorème de Markov

On peut facilement généraliser le théorème de Tchébychev au cas plus compliqué où la loi de répartition de la variable aléatoire X change d'une expérience à l'autre. Dans ce cas, au lieu de la moyenne arithmétique des valeurs observées d'une même grandeur X d'espérance mathématique et de variance constantes, nous avons affaire à la moyenne arithmétique de n variables aléatoires différentes, d'espérances mathématiques et de variances différentes. Il se trouve que même dans ce cas, sous certaines réserves, la moyenne arithmétique est stable et converge en probabilité vers une certaine grandeur non aléatoire.

Le théorème de Tchébychev généralisé peut se formuler comme suit. *Si*

$$X_1, X_2, \ldots, X_n$$

sont des variables aléatoires indépendantes d'espérances mathématiques

$$m_{x_1}, m_{x_2}, \ldots, m_{x_n}$$

et de variances

$$D_{x_1}, D_{x_2}, \ldots, D_{x_n},$$

et si toutes les variances ont même limite supérieure L:

$$D_{x_i} < L \qquad (i = 1, 2, \ldots, n),$$

lorsque n augmente la moyenne arithmétique des valeurs observées des variables $X_1, X_2, \ldots, X_n$ converge en probabilité vers la moyenne arithmétique de leurs espérances mathématiques *).

Nous allons écrire ce théorème sous la forme d'une formule. Soient ε, δ des nombres positifs aussi petits que l'on veut. Pour n suffisamment grand on aura alors:

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{x_i}}{n}\right| < \varepsilon\right) > 1 - \delta. \qquad (13.4.1)$$

D é m o n s t r a t i o n. Considérons la grandeur:

$$Y = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Son espérance mathématique est:

$$m_y = \frac{\sum_{i=1}^{n} m_{x_i}}{n},$$

et sa variance:

$$D_y = \frac{1}{n^2} \sum_{i=1}^{n} D_{x_i}.$$

Appliquons à Y l'inégalité de Tchébychev, soit:

$$P(|Y - m_y| \geqslant \varepsilon) \leqslant \frac{D_y}{\varepsilon^2},$$

ou

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{x_i}}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\sum_{i=1}^{n} D_{x_i}}{n^2 \varepsilon^2}. \qquad (13.4.2)$$

*) C'est-à-dire que la différence entre les deux moyennes arithmétiques converge en probabilité vers zéro.

Remplaçons dans le second membre de (13.4.2) chacune des grandeurs D_{x_i} par la grandeur L qui lui est supérieure. Ceci ne fera que renforcer l'inégalité :

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n}-\frac{\sum_{i=1}^{n} m_{x_i}}{n}\right|\geqslant \varepsilon\right)<\frac{lL}{n\varepsilon^2}.$$

Quelque petit que soit ε on peut toujours trouver un n suffisamment grand pour que l'inégalité suivante soit remplie

$$\frac{L}{n\varepsilon^2}<\delta;$$

on a alors :

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n}-\frac{\sum_{i=1}^{n} m_{x_i}}{n}\right|\geqslant \varepsilon\right)<\delta,$$

d'où, en passant à l'événement contraire, on obtient l'inégalité à démontrer (13.4.1).

On peut généraliser la loi des grands nombres au cas des variables aléatoires dépendantes. On doit cette généralisation à A. Markov.

Théorème de Markov. *Soient* $X_1, X_2, \ldots, X_n$ *des variables aléatoires dépendantes. Si pour* $n \to \infty$ *on a* :

$$\frac{D\left[\sum_{i=1}^{n} X_i\right]}{n^2}\to 0,$$

la moyenne arithmétique des valeurs observées des variables aléatoires $X_1, X_2, \ldots, X_n$ *converge en probabilité vers la moyenne arithmétique de leurs espérances mathématiques.*

Démonstration. Considérons la grandeur :

$$Y=\frac{\sum_{i=1}^{n} X_i}{n}.$$

Il est évident que l'on a :

$$D_y=\frac{D\left[\sum_{i=1}^{n} X_i\right]}{n^2}.$$

Appliquons à Y l'inégalité de Tchébychev :

$$P(|Y-m_y|\geqslant\varepsilon)\leqslant\frac{D_y}{\varepsilon^2}.$$

Comme en vertu de ce théorème pour $n \to \infty$ on a $D_y \to 0$, pour n suffisamment grand :

$$P(|Y - m_y| \geqslant \varepsilon) < \delta,$$

ou, en passant à l'événement contraire :

$$P(|Y - m_y| < \varepsilon) = P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{x_i}}{n}\right| < \varepsilon\right) > 1 - \delta,$$

ce qu'il fallait démontrer.

13.5. Conséquences de la loi des grands nombres : théorèmes de Bernoulli et de Poisson

Le théorème bien connu de J. Bernoulli établissant une relation entre la fréquence d'un événement et sa probabilité peut être démontré comme une conséquence immédiate de la loi des grands nombres.

Supposons que l'on effectue n expériences indépendantes, dans chacune desquelles l'événement A peut apparaître ou ne pas apparaître, avec une probabilité p. En vertu du théorème de J. Bernoulli, *lorsque le nombre d'expériences n augmente indéfiniment, la fréquence de l'événement A converge en probabilité vers sa probabilité p.*

Désignons par P^* la fréquence de l'événement A dans n expériences, le théorème de J. Bernoulli peut alors s'écrire comme suit :

$$P(|P^* - p| < \varepsilon) > 1 - \delta, \qquad (13.5.1)$$

où ε, δ sont des nombres positifs aussi petits que l'on veut.

Il y a lieu de démontrer cette formule pour un n suffisamment grand.

Démonstration. Considérons les variables aléatoires indépendantes : X_1, nombre d'apparitions de l'événement A dans la première expérience; X_2, nombre d'apparitions de l'événement A dans la seconde expérience, etc.

Toutes ces variables sont discontinues et ont même loi de répartition s'exprimant par le tableau de la forme :

0	1
q	p

avec $q = 1 - p$. L'espérance mathématique de chacune des variables X_i est égale à p, et sa variance à pq (voir § 10.3).

La fréquence P^* n'est rien d'autre que la moyenne arithmétique des variables $X_1, X_2, \ldots, X_n$:

$$P^* = \frac{\sum_{i=1}^{n} X_i}{n}$$

et en vertu de la loi des grands nombres converge en probabilité vers l'espérance mathématique commune à ces variables aléatoires, d'où l'inégalité (13.5.1).

En vertu du théorème de Bernoulli pour des conditions d'expériences constantes la fréquence est stable. Cette stabilité se manifeste également lorsque les conditions d'expériences sont variables. Le théorème établissant la stabilité des fréquences pour des conditions d'expériences variables est appelé *théorème de Poisson* et se formule comme suit:

Si l'on effectue n expériences indépendantes et que la probabilité d'apparition de l'événement A dans l' i-ème expérience est égale à p_i*, avec augmentation de n, la fréquence de l'événement A converge en probabilité vers la moyenne arithmétique des* p_i.

Le théorème de Poisson peut être déduit du théorème généralisé de Tchébychev tout comme le théorème de Bernoulli a été déduit à partir de la loi des grands nombres.

Le théorème de Poisson joue un très grand rôle dans les applications pratiques de la théorie des probabilités. En effet, les méthodes probabilistes sont utilisées pour l'étude des effets qui ont peu de chances de se répéter dans les mêmes conditions un grand nombre de fois, mais se répètent maintes fois dans des conditions très différentes, la probabilité des événements nous intéressant dépendant fortement de ces conditions. Par exemple, la probabilité d'atteindre un objectif dépend beaucoup de la distance de l'objectif, de raccourci de l'objectif, de l'altitude et de la vitesse de l'avion et de l'objectif, etc. Les facteurs déterminant les conditions du combat aérien sont trop nombreux pour qu'on puisse s'attendre à la répétition des mêmes conditions. Néanmoins, on peut noter une stabilité de la fréquence d'atteindre un but dans des expériences réelles pour des conditions très variables, cette fréquence tend vers la probabilité moyenne d'atteindre un but caractéristique pour un groupe donné de conditions. C'est pourquoi les méthodes d'organisation de tirs, basées sur la probabilité maximale d'atteindre le but, seront ici justifiées, bien que les expériences ne soient pas nombreuses pour un ensemble donné de conditions.

La situation est analogue quand il s'agit de vérifier expérimentalement les calculs probabilistes. Souvent il y a lieu de vérifier, que la probabilité calculée d'un événement quelconque A correspond à sa fréquence donnée. Le plus souvent il s'agit de vérifier un schéma théorique se trouvant à la base d'une méthode de calcul de la proba-

bilité d'un événement. Lors d'une telle vérification on n'arrive pas toujours à reproduire les mêmes conditions d'expériences un nombre suffisant de fois. Néanmoins on peut réaliser cette vérification si l'on compare la fréquence observée non pas avec la probabilité pour des conditions fixées mais avec la *moyenne arithmétique* des probabilités calculées pour des conditions différentes.

13.6. Les phénomènes aléatoires de masse et le théorème central limite

Dans les paragraphes précédents nous avons étudié différentes formes de la loi des grands nombres, affirmant toutes la même chose, à savoir la convergence en probalilité des variables aléatoires vers des grandeurs constantes. Dans aucune forme de la loi des grands nombres nous n'avons envisagé les lois de répartition des variables aléatoires. Les lois de répartition limites sont objet d'un autre groupe de théorèmes, appelé theorème central limite, parfois considéré comme forme quantitative de la loi des grands nombres.

Toutes les formes du théorème central limite établissent les conditions pour lesquelles une loi normale peut apparaître. Comme ces conditions sont assez souvent réalisées dans la pratique, la loi normale se trouve être la plus répandue des lois de répartition, surtout dans les phénomènes aléatoires de la nature. Elle apparaît chaque fois que la grandeur aléatoire observée peut être présentée comme la somme d'un nombre suffisamment grand de composantes élémentaires indépendantes (ou faiblement liées), dont chacune influe peu sur la somme.

Dans la théorie du tir la loi de répartition normale joue un rôle particulièrement important, car dans la plupart des cas pratiques les coordonnées des points d'atteinte et des points d'éclatement des obus sont normalement réparties. Ceci peut s'expliquer sur l'exemple suivant.

Supposons qu'on tire sur une cible plane dont le centre (point de visée) est pris pour origine des coordonnées. Le point d'atteinte peut être caractérisé par deux variables aléatoires X et Y. Considérons l'une d'elles, par exemple l'écart X entre le point d'atteinte et le but suivant l'axe Ox. Cet écart est dû à l'action conjointe d'un grand nombre de facteurs relativement peu importants, comme, par exemple, l'erreur de pointage, l'erreur de détermination de la distance à la cible, les vibrations de la pièce lors du tir, les erreurs de fabrication de l'obus, les conditions atmosphériques, etc. Chacun de ces facteurs produit une erreur X_i, dite élémentaire, dont la somme donne la coordonnée X du point d'atteinte:

$$X = X_1 + X_2 + \ldots + X_n + \ldots \qquad (13.6.1)$$

Comme ces facteurs sont très nombreux, ils ne dépendent presque pas les uns des autres et l'influence de chacun sur la somme est petite et

à peu près la même, les conditions d'applicabilité du théorème central limite se trouvent ainsi réalisées et la grandeur (13.6.1) doit obéir à une loi de répartition voisine de la loi normale.

Voyons un peu plus en détail l'assertion sur l'influence petite et à peu près égale de chacun des termes sur la somme. Le sens en est que parmi les erreurs élémentaires de tir aucune ne prévaut sur la somme de toutes les autres. En effet, si une erreur prédominait, il y a lieu de penser qu'en établissant les lois de tir ou en construisant un appareil de visée, on essayerait de l'éviter en tenant compte de la plus importante des causes écartant l'obus du but. Les facteurs aléatoires négligés donnant lieu à une dispersion sont généralement très petits et aucun d'eux ne prédomine. C'est pourquoi on adopte généralement la loi normale pour la répartition des points d'impact des obus *).

La loi de répartition normale prédomine non seulement dans la théorie du tir mais également dans de nombreux autres domaines, par exemple dans la théorie des mesures. C'est en se basant sur la théorie des mesures que Laplace et Gauss ont formulé la loi normale. En effet, dans la majorité des cas les erreurs apparaissant lors de la mesure de telle ou telle grandeur physique obéissent à la loi normale ; la raison en est que ces erreurs sont généralement la somme d'un grand nombre d'erreurs élémentaires indépendantes dues à des causes différentes. Longtemps la loi normale était considérée comme loi unique et universelle des erreurs. Ce n'est pas vrai, car l'expérience montre que dans certains processus de mesure et de fabrication on rencontre des lois de répartition différant de la loi normale, néanmoins celle-ci reste la plus répandue et la plus importante pour les applications pratiques des lois de mesure.

13.7. Fonctions caractéristiques

Dans sa forme générale le théorème central limite a été démontré par A. Liapounov en 1900. Pour démontrer ce théorème A. Liapounov a élaboré une méthode spéciale dite des *fonctions caractéristiques*. Par la suite cette méthode a acquis une importance par elle-même, s'étant avérée puissante et souple.

On appelle *fonction caractéristique* de la variable aléatoire X la fonction

$$g(t) = M[e^{itX}], \tag{13.7.1}$$

*) Dans certains cas la répartition réelle des points d'impact dans le plan peut différer notablement de la loi normale, par exemple lors du tir dans des conditions très variables, lorsque le centre de dispersion et l'écart probable lors du tir changent notablement. Cependant nous avons alors affaire non pas à la loi de répartition des coordonnées du point d'impact pour un coup, mais à la moyenne de ces lois pour des coups différents.

où i est l'unité imaginaire. La fonction $g(t)$ est l'espérance mathématique d'une certaine variable aléatoire complexe

$$U = e^{itX},$$

fonctionnellement liée à X. Pour la solution de nombreux problèmes de la théorie des probabilités il est plus commode d'utiliser les fonctions caractéristiques et non pas les lois de répartition.

Connaissant la loi de répartition de X il est facile de trouver sa fonction caractéristique.

Si X est une variable aléatoire discrète répartie suivant le tableau :

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

sa fonction caractéristique est:

$$g(t) = \sum_{k=1}^{n} e^{itx_k} p_k. \qquad (13.7.2)$$

Si X est une variable aléatoire continue de densité de probabilité $f(x)$, sa fonction caractéristique est:

$$g(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx. \qquad (13.7.3)$$

Exemple 1. La variable aléatoire X est le nombre d'impacts lors d'un coup. La probabilité d'atteindre le but est égale à p. Trouver la fonction caractéristique de X.

Solution. En vertu de la formule (13.7.2) on a:

$$g(t) = e^{it \cdot 0}(1-p) + e^{it \cdot 1} p = q + e^{it} p$$

où $q = 1 - p$.

Exemple 2. La variable aléatoire X est répartie suivant la loi normale:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \qquad (13.7.4)$$

Trouver sa fonction caractéristique.

Solution. En vertu de la formule (13.7.3) on a:

$$g(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} dx. \qquad (13.7.5)$$

En utilisant la formule bien connue

$$\int_{-\infty}^{\infty} e^{-Ax^2 \pm 2Bx - C} dx = \sqrt{\frac{\pi}{A}}\, e^{-\frac{AC-B^2}{A}}$$

et compte tenu de ce que $i^2 = -1$ on a:

$$g(t) = e^{-\frac{t^2}{2}}. \qquad (13.7.6)$$

La formule (13.7.3) donne la fonction caractéristique $g(t)$ d'une variable aléatoire continue X en fonction de sa densité de probabilité $f(x)$. La transformation (13.7.3) permettant de trouver $g(t)$ à partir de $f(x)$ porte le nom de *transformation de Fourier*. Dans les cours d'analyse mathématique on montre que si la fonction $g(t)$ s'exprime en fonction de $f(x)$ à l'aide d'une transformation de Fourier, inversement, la fonction $f(x)$ s'exprime en fonction de $g(t)$ à l'aide de la *transformation de Fourier inverse*:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} g(t)\, dt\ ^{*}). \qquad (13.7.7)$$

Nous allons formuler et démontrer les propriétés essentielles des fonctions caractéristiques.

1. *Si deux variables aléatoires X et Y sont liées par la relation*

$$Y = aX,$$

où a est un facteur non aléatoire, leurs fonctions caractéristiques sont liées par la relation

$$g_y(t) = g_x(at). \qquad (13.7.8)$$

Démonstration.

$$g_y(t) = M[e^{itY}] = M[e^{itaX}] = M[e^{i(at)X}] = g_x(at).$$

2. *La fonction caractéristique d'une somme de variables aléatoires indépendantes est égale au produit des fonctions caractéristiques des composantes.*

Démonstration. Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes de fonctions caractéristiques

$$g_{x_1}(t),\ g_{x_2}(t),\ \ldots,\ g_{x_n}(t)$$

et leur somme:

$$Y = \sum_{k=1}^{n} X_k.$$

Il y a lieu de montrer que:

$$g_y(t) = \prod_{k=1}^{n} g_{x_k}(t). \qquad (13.7.9)$$

*) Aux §§ 17.3 et 17.4 on verra les mêmes transformations de Fourier mais reliant la fonction de corrélation et la densité spectrale.

On a:

$$g_y(t) = M[e^{itY}] = M\left[e^{it\sum_{k=1}^{n} X_k}\right] = M[\prod_{k=1}^{n} e^{itX_k}].$$

Les variables X_k étant indépendantes, leurs fonctions e^{itX_k} sont également indépendantes. En vertu du théorème du produit des espérances mathématiques on a:

$$g_y(t) = \prod_{k=1}^{n} M[e^{itX_k}] = \prod_{k=1}^{n} g_{x_k}(t),$$

ce qu'il fallait démontrer.

L'appareil des fonctions caractéristiques est souvent utilisé pour la composition des lois de répartition. Soient par exemple deux variables aléatoires indépendantes X et Y de densités de probabilité $f_1(x)$ et $f_2(y)$. Il y a lieu de trouver la densité de probabilité de la variable

$$Z = X + Y.$$

A cet effet on peut procéder comme suit: on trouve les fonctions caractéristiques $g_x(t)$ et $g_y(t)$ des variables aléatoires X et Y, on forme leur produit et on obtient ainsi la fonction caractéristique de la grandeur Z:

$$g_z(t) = g_x(t)\, g_y(t),$$

puis en faisant subir à $g_z(t)$ une transformation de Fourier inverse on trouve la densité de probabilité de Z:

$$f_3(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz} g_z(t)\, dt.$$

Exemple 3. Trouver à l'aide des fonctions caractéristiques la composition de deux lois normales:

$f_1(x)$ de caractéristiques $m_x = 0$; σ_x;

$f_2(x)$ de caractéristiques $m_y = 0$; σ_y.

Solution. Cherchons la fonction caractéristique de X. A cet effet écrivons:

$$X = \sigma_x U,$$

où $m_u = 0$; $\sigma_u = 1$.

En utilisant le résultat de l'exemple 2 on trouve:

$$g_u(t) = e^{-\frac{t^2}{2}}.$$

En vertu de la propriété 1 des fonctions caractéristiques on a:

$$g_x(t) = g_u(\sigma_x t) = e^{-\frac{(\sigma_x t)^2}{2}}$$

et d'une manière analogue:

$$g_y(t) = e^{-\frac{(\sigma_y t)^2}{2}}.$$

En multipliant $g_x(t)$ par $g_y(t)$ on a:

$$g_z(t) = e^{-\frac{(\sigma_x^2+\sigma_y^2)}{2}t^2},$$

c'est la fonction caractéristique d'une loi normale de paramètres $m_z = 0$; $\sigma_z = \sqrt{\sigma_x^2+\sigma_y^2}$. On a ainsi obtenu la composition des lois normales d'une manière bien plus simple que dans le § 12.6.

13.8. Théorème central limite pour des composantes de même loi de répartition

Les différentes formes du théorème central limite diffèrent les unes des autres par les conditions imposées à la répartition des composantes aléatoires. Nous allons formuler et démontrer ici le théorème central limite pour le cas le plus simple où les termes sont de même loi de répartition.

Théorème. *Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes de même loi de répartition, d'espérance mathématique m et de variance σ^2; lorsque n augmente indéfiniment la loi de répartition de la somme*

$$Y_n = \sum_{k=1}^{n} X_k \tag{13.8.1}$$

tend vers la loi normale.

Démonstration. Nous allons donner la démonstration pour le cas des variables aléatoires continues $X_1, \ldots, X_n$ (dans le cas discret la démonstration est analogue).

En vertu de la seconde propriété des fonctions caractéristiques démontrée dans le paragraphe précédent, la fonction caractéristique de Y_n est égale au produit des fonctions caractéristiques des composantes. Les variables aléatoires $X_1, \ldots, X_n$ ont même loi de répartition de densité $f(x)$ et par conséquent même fonction caractéristique

$$g_x(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx. \tag{13.8.2}$$

La fonction caractéristique de Y_n sera:

$$g_{y_n}(t) = [g_x(t)]^n. \tag{13.8.3}$$

Nous allons étudier plus en détail la fonction $g_x(t)$. A cet effet écrivons-la au voisinage du point $t = 0$ comme une série de Maclaurin,

en nous bornant aux trois termes:

$$g_x(t) = g_x(0) + g'_x(0)\,t + \left[\frac{g''_x(0)}{2} + \alpha(t)\right] t^2, \qquad (13.8.4)$$

où $\alpha(t) \to 0$ pour $t \to 0$.

Calculons les grandeurs $g_x(0)$, $g'_x(0)$, $g''_x(0)$. En posant $t = 0$ dans la formule (13.8.2) on a:

$$g_x(0) = \int_{-\infty}^{\infty} f(x)\,dx = 1. \qquad (13.8.5)$$

Dérivons (13.8.2) par rapport à t:

$$g'_x(t) = \int_{-\infty}^{\infty} ix\,e^{itx} f(x)\,dx = i \int_{-\infty}^{\infty} x e^{itx} f(x)\,dx. \qquad (13.8.6)$$

En posant $t = 0$ dans (13.8.6) on obtient:

$$g'_x(0) = i \int_{-\infty}^{\infty} x f(x)\,dx = iM[X] = im. \qquad (13.8.7)$$

Il est évident qu'on ne restreint pas la généralité en posant $m = 0$ (il suffit pour cela de transférer l'origine au point m). On a alors:

$$g'_x(0) = 0.$$

En dérivant (13.8.6) une deuxième fois on a:

$$g''_x(t) = -\int_{-\infty}^{\infty} x^2 e^{itx} f(x)\,dx,$$

d'où

$$g''_x(0) = -\int_{-\infty}^{\infty} x^2 f(x)\,dx. \qquad (13.8.8)$$

Pour $m = 0$ l'intégrale dans l'expression (13.8.8) n'est rien d'autre que la variance de X de densité $f(x)$, par conséquent:

$$g''_x(0) = -\sigma^2. \qquad (13.8.9)$$

En substituant $g_x(0) = 1$, $g'_x(0) = 0$ et $g''_x(0) = -\sigma^2$ dans (13.8.4) on obtient:

$$g_x(t) = 1 - \left[\frac{\sigma^2}{2} - \alpha(t)\right] t^2. \qquad (13.8.10)$$

Considérons la variable Y_n. Nous voulons démontrer que sa loi de répartition lorsque n augmente tend vers la loi normale. Pour cela

passons de Y_n à la variable aléatoire normée

$$Z_n = \frac{Y_n}{\sigma\sqrt{n}} \tag{13.8.11}$$

très commode, car sa variance ne dépend pas de n et est égale à l'unité quel que soit n. Il est facile de s'en rendre compte en considérant Z_n comme une fonction linéaire des variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$ de même variance σ^2. Si l'on montre que la loi de répartition de Z_n tend vers la loi normale, il est évident qu'il en sera de même pour Y_n, liée à Z_n par la relation linéaire (13.8.11).

Au lieu de démontrer la convergence de la loi de Z_n, avec augmentation de n, vers la loi normale, nous allons démontrer que sa fonction caractéristique tend vers la fonction caractéristique de la loi normale *).

Cherchons la fonction caractéristique de Z_n. En vertu de la première propriété (13.7.8) des fonctions caractéristiques on obtient à partir de la relation (13.8.11):

$$g_{z_n}(t) = g_{y_n}\left(\frac{t}{\sigma\sqrt{n}}\right), \tag{13.8.12}$$

où $g_{y_n}(t)$ est la fonction caractéristique de Y_n.

Les formules (13.8.12) et (13.8.3) donnent:

$$g_{z_n}(t) = \left[g_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \tag{13.8.13}$$

ou, en utilisant la formule (13.8.10):

$$g_{z_n}(t) = \left\{1 - \left[\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\frac{t^2}{n\sigma^2}\right\}^n. \tag{13.8.14}$$

Prenons le logarithme de l'expression (13.8.14):

$$\ln g_{z_n}(t) = n \ln\left\{1 - \left[\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\frac{t^2}{n\sigma^2}\right\}.$$

Introduisons la désignation:

$$\left[\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\frac{t^2}{n\sigma^2} = \varkappa. \tag{13.8.15}$$

On a alors:

$$\ln g_{z_n}(t) = n \ln\{1 - \varkappa\}. \tag{13.8.16}$$

Faisons tendre n vers l'infini. En vertu de (13.8.15) ceci fait tendre $\varkappa$ vers zéro. Lorsque n est important, cette grandeur est assez petite. Développons $\ln\{1 - \varkappa\}$ en série et limitons-nous au premier terme du développement (pour $n \to \infty$ les autres termes seront négli-

*) Nous adoptons ici sans démonstration que la convergence des fonctions caractéristiques entraîne la convergence en lois.

geables) :

$$\ln\{1-x\}\approx -x.$$

On aura alors :

$$\lim_{n\to\infty}\ln g_{z_n}(t)=\lim_{n\to\infty} n\cdot(-x)=$$

$$=\lim_{n\to\infty}\left\{-\frac{t^2}{2}+\alpha\left(\frac{t}{\sigma\sqrt{n}}\right)\frac{t^2}{\sigma^2}\right\}=-\frac{t^2}{2}+\lim_{n\to\infty}\frac{t^2}{\sigma^2}\alpha\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Par définition la fonction $\alpha(t)$ tend vers zéro pour $t\to 0$; par conséquent :

$$\lim_{n\to\infty}\alpha\left(\frac{t}{\sigma\sqrt{n}}\right)=0$$

et

$$\lim_{n\to\infty}\ln g_{z_n}(t)=-\frac{t^2}{2},$$

d'où

$$\lim_{n\to\infty} g_{z_n}(t)=e^{-\frac{t^2}{2}}. \qquad (13.8.17)$$

Ce n'est rien d'autre que la fonction caractéristique de la loi normale de paramètres $m=0$, $\sigma=1$ (voir exemple 2 du § 13.7).

Ainsi on a démontré que lorsque n augmente la fonction caractéristique de la variable aléatoire Z_n tend indéfiniment vers la fonction caractéristique de la loi normale ; on en conclut que la loi de Z_n (donc également de Y_n) converge vers la loi normale, ce qu'il fallait prouver.

Nous avons ainsi démontré le théorème central limite pour le cas particulier, mais fort important, de composantes de même répartition. Cependant ce théorème est également vrai, sans restrictions importantes, pour des composantes à répartition différente. A. Liapounov a démontré le théorème central limite pour les conditions suivantes :

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n} b_k}{\left\{\sum_{k=1}^{n} D_k\right\}^{3/2}}=0, \qquad (13.8.18)$$

où b_k est le troisième moment central absolu de X_k :

$$b_k=\nu_3[X_k]=M[|\overset{\circ}{X}_k|^3] \quad (k=1,\ \ldots,\ n)$$

et D_k, la variance de X_k.

La condition la plus générale (nécessaire et suffisante) pour que le théorème central limite soit vrai est la *condition de Lindeberg* : *pour*

$\tau > 0$ *quelconque*

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-m_k|>\tau B_n} (x-m_k)^2 f_k(x)\, dx = 0,$$

où m_k *est l'espérance mathématique,* $f_k(x)$ *la densité de probabilité de la grandeur aléatoire* X_k *et* $B_n = \sqrt{\sum_{k=1}^{n} D_k}$.

13.9. Formules pratiques exprimant le théorème central limite

En vertu du théorème central limite, la loi de répartition d'une somme d'un nombre suffisamment grand de variables aléatoires indépendantes (pour des restrictions assez faibles) converge vers la loi normale. On peut utiliser le théorème central limite même lorsqu'on a affaire à une somme d'un nombre relativement peu important de variables aléatoires. Lorsque l'on prend la somme de variables aléatoires indépendantes, de dispersion voisine, au fur et à mesure de l'augmentation du nombre des composantes la loi de répartition de la somme converge rapidement vers la loi normale. Dans la pratique on remplace souvent une loi de répartition par une autre ; pour une précision assez faible des calculs probabilistes, cette substitution peut également être très approchée. L'expérience montre, que pour une dizaine de composantes (parfois même moins) la loi de répartition de la somme peut être remplacée par la loi normale.

On utilise souvent le théorème central limite pour calculer la probabilité pour la somme de plusieurs variables aléatoires de se trouver dans certaines limites.

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes d'espérances mathématiques

$$m_1, m_2, \ldots, m_n$$

et de variances

$$D_1, D_2, \ldots, D_n.$$

Supposons que les conditions du théorème central limite soient remplies : les variables $X_1, X_2, \ldots, X_n$ produisent les effets comparables sur la dispersion de la somme et le nombre de termes n est suffisant pour que la loi de répartition de la somme

$$Y = \sum_{i=1}^{n} X_i \tag{13.9.1}$$

puisse être considérée normale.

La probabilité pour la variable aléatoire Y de se trouver dans l'intervalle (α, β) s'exprime par la formule:

$$P(\alpha < Y < \beta) = \Phi^*\left(\frac{\beta - m_y}{\sigma_y}\right) - \Phi^*\left(\frac{\alpha - m_y}{\sigma_y}\right), \quad (13.9.2)$$

où m_y, σ_y sont l'espérance mathématique et l'écart quadratique moyen de Y; Φ^* la fonction de répartition normale.

En vertu du théorème d'addition des espérances mathématiques et des variances on a:

$$\left.\begin{aligned} m_y &= \sum_{i=1}^{n} m_i, \\ \sigma_y &= \sqrt{D_y} = \sqrt{\sum_{i=1}^{n} D_i}. \end{aligned}\right\} \quad (13.9.3)$$

Ainsi pour trouver approximativement la probabilité pour la somme d'un grand nombre de variables aléatoires de se trouver à l'intérieur d'un certain intervalle, il n'est pas nécessaire de connaître la loi de répartition de ces variables mais seulement leurs caractéristiques. Il est évident que ceci n'est vrai que si se trouve remplie la condition essentielle du théorème central limite, à savoir que chaque terme exerce une influence petite et du même ordre de grandeur sur la dispersion de la somme.

En plus des formules du type (13.9.2), dans les applications pratiques on utilise des formules dans lesquelles la somme des X_i est remplacée par leur somme normée, soit:

$$Z = \frac{\overset{\circ}{Y}}{\sigma_y} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} D_i}}. \quad (13.9.4)$$

Il est évident que l'on a:

$$M[Z] = 0; \quad D[Z] = \sigma_z = 1.$$

Si la loi de répartition de Y est voisine de la loi normale de paramètres (13.9.3), celle de Z est voisine d'une loi normale de paramètres $m_z = 0$, $\sigma_z = 1$. D'où

$$P(\alpha < Z < \beta) = \Phi^*(\beta) - \Phi^*(\alpha). \quad (13.9.5)$$

Le théorème central limite peut être appliqué non seulement aux variables aléatoires continues, mais également aux variables aléatoires discrètes à condition d'opérer non pas avec les densités mais avec les fonctions de répartition. En effet, si les variables $X_1, X_2, \ldots, X_n$ sont discrètes, leur somme est également une variable aléatoire discrète, et, en toute rigueur, ne peut suivre une loi normale.

Néanmoins, toutes les formules du type (13.9.2), (13.9.5) restent vraies car elles contiennent non pas des densités mais des fonctions de répartition. On peut montrer que si les variables aléatoires discrètes satisfont au théorème central limite, la fonction de répartition de leur somme normée Z [voir la formule (13.9.4)] pour $n \to \infty$ converge vers une loi normale de paramètres $m_z = 0$, $\sigma_z = 1$.

Le *théorème de Laplace* est un cas particulier du théorème central limite pour des variables aléatoires discrètes.

Si l'on effectue n expériences indépendantes, dans chacune desquelles l'événement A apparaît avec une probabilité p on a la relation :

$$P\left(\alpha < \frac{Y-np}{\sqrt{npq}} < \beta\right) = \Phi^*(\beta) - \Phi^*(\alpha), \qquad (13.9.6)$$

où Y est le nombre d'apparitions de l'événement A dans n expériences et $q = 1 - p$.

Démonstration. Supposons que l'on effectue n expériences indépendantes dans chacune desquelles l'événement A peut apparaître avec une probabilité p. Ecrivons Y, le nombre total d'apparitions de l'événement A dans n expériences, comme la somme :

$$Y = \sum_{i=1}^{n} X_i, \qquad (13.9.7)$$

où X_i est le nombre d'apparitions de l'événement A dans l'i-ème expérience.

En vertu du théorème démontré dans le § 13.8, la loi de répartition d'une somme de termes de même répartition tend vers la loi normale lorsque le nombre de termes augmente. Par conséquent, pour un n suffisamment grand on peut appliquer la formule (13.9.5) avec

$$Z = \frac{Y - m_y}{\sigma_y}. \qquad (13.9.8)$$

Dans le § 10.3 nous avons démontré que l'espérance mathématique et la variance du nombre d'apparitions de l'événement dans n expériences indépendantes sont :

$$m_y = np; \quad D_y = npq \; (q = 1 - p).$$

En substituant ces expressions dans (13.9.8) on obtient :

$$Z = \frac{Y-np}{\sqrt{npq}},$$

la formule (13.9.5) devient alors :

$$P\left(\alpha < \frac{Y-np}{\sqrt{npq}} < \beta\right) = \Phi^*(\beta) - \Phi^*(\alpha),$$

ce qui démontre le théorème énoncé.

E x e m p l e 1. Supposons que 120 personnes fassent la queue à la caisse pour recevoir leurs paies, primes et assurances maladie. L'espérance mathématique de la somme payée à une personne est égale à 50 francs, l'écart quadratique moyen est 30 francs. La caisse dispose de 6 500 francs. Trouver la probabilité pour que cette somme d'argent suffise pour payer les personnes de la queue.

S o l u t i o n. Considérons la variable aléatoire X, égale à la somme que doit payer la caisse. C'est une somme de $n = 120$ variables aléatoires indépendantes; le nombre de termes est grand; en vertu du théorème central limite la loi de X est voisine d'une loi normale. Les caractéristiques de cette loi sont: $m_x = 120 \cdot 50 = 6000$, la variance $D_x = 120 \cdot 30^2 = 108\,000$, l'écart quadratique moyen $\sigma_x = \sqrt{108\,000} \approx 328$. La probabilité pour que la somme disponible soit suffisante pour payer tout le monde est:

$$P(X \leqslant 6500) \approx P(X < 6500) = \Phi^*\left(\frac{6500-6000}{328}\right) \approx 0{,}936.$$

E x e m p l e 2. Des appareils de deux types I et II, de même destination sont en service. Un appareil du type I doit être réparé après exploitation avec la probabilité $p_1 = 0{,}2$, et un appareil du type II avec la probabilité $p_2 = 0{,}4$. Le nombre total d'appareils en service est $n_1 = 1000$ du type I et $n_2 = 1500$ du type II. 1) Trouver la probabilité pour que le nombre d'appareils des deux types devant être réparés se trouve entre 750 et 850. 2) Trouver le nombre de réparations que doivent envisager les ateliers pour qu'avec une probabilité égale 0,99 le nombre d'appareils à réparer ne soit pas supérieur à N.

S o l u t i o n. Lorsque le nombre d'expériences est important (ici on a $n_1 + n_2 = 2\,500$) on peut considérer que le nombre X d'apparitions de l'événement A — l'appareil doit être réparé — suit une loi normale. Les caractéristiques de la variable aléatoire X sont: $m_x = 1000 \cdot 0{,}2 + 1\,500 \cdot 0{,}4 = 200 + 600 = 800$; $D_x = 1000 \cdot 0{,}2 \cdot 0{,}8 + 1\,500 \cdot 0{,}4 \cdot 0{,}6 = 520$, $\sigma_x = \sqrt{520} \approx 22{,}8$.

1) $P(750 \leqslant X \leqslant 850) \approx P(750 < X < 850) = P(|X - 800| < 50) = 2\Phi^*\left(\frac{50}{22{,}8}\right) - 1 = 2\Phi^*(2{,}19) - 1 \approx 0{,}972$.

2) La valeur x_0 qui, avec une probabilité $p = 0{,}99$, ne sera pas dépassée par le nombre d'appareils envoyés en réparation, est donnée par la formule:

$$0{,}99 = P(X < x_0) = \Phi^*\left(\frac{x_0 - 800}{22{,}8}\right).$$

Dans la table 1 de l'annexe on peut trouver la valeur de l'argument pour lequel la fonction Φ^* est égale à 0,99: $\Phi^*(2{,}33) \approx 0{,}99$; d'où $\frac{x_0 - 800}{22{,}8} = 2{,}33$ et $x_0 \approx 853$.

CHAPITRE 14

TRAITEMENT DES RÉSULTATS DES EXPÉRIENCES

14.1. Particularités du traitement des résultats d'un nombre limité d'expériences. Estimations des paramètres inconnus d'une loi de répartition

Dans le chapitre 7 nous avons déjà étudié certains problèmes de mathématiques statistiques relatifs au traitement des résultats des expériences. Il s'agissait essentiellement de trouver les lois de répartition de grandeurs aléatoires d'après les résultats des expériences. Pour pouvoir déterminer la loi de répartition, il faut disposer de données statistiques suffisantes, plusieurs centaines au moins. Cependant dans la pratique, on a affaire à un nombre de données statistiques bien plus réduit : deux ou trois dizaines d'observations, souvent même moins. Cet état de choses s'explique par le prix élevé et par la complexité des expériences. Des données aussi limitées sont nettement insuffisantes pour trouver la loi de répartition inconnue d'une variable aléatoire; néanmoins ces données peuvent être traitées et utilisées pour obtenir certains renseignements sur la variable aléatoire étudiée. Par exemple, on peut, à partir de données statistiques limitées, déterminer, ne serait-ce qu'approximativement, des caractéristiques numériques importantes d'une variable aléatoire, à savoir l'espérance mathématique, la variance et parfois les moments d'ordre plus élevé. Dans la pratique on connaît souvent à l'avance la forme de la loi de répartition et il ne reste qu'à déterminer certains paramètres de cette dernière. Si, par exemple, on sait qu'une variable aléatoire est normale, le traitement se réduit à la détermination de deux paramètres de cette loi : m et σ. Si l'on sait que la variable aléatoire étudiée est répartie suivant la loi de Poisson, un seul paramètre est à déterminer, à savoir l'espérance mathématique a. Enfin, dans certains problèmes la forme de la loi de répartition importe peu, seules les caractéristiques numériques doivent être déterminées.

Dans le présent chapitre nous allons étudier un certain nombre de problèmes où il y a lieu de déterminer les paramètres inconnus de la loi de répartition d'une variable aléatoire, ceci à partir d'un nombre limité de données.

Avant tout il faut noter que la valeur d'un paramètre inconnu, calculée à partir d'un nombre restreint de données, aura toujours un caractère aléatoire. Nous appellerons *estimation* d'un paramètre cette valeur aléatoire approchée. Par exemple, l'estimation de l'espérance

mathématique peut être la moyenne arithmétique des valeurs observées d'une variable aléatoire lors de n expériences indépendantes. Lorsque le nombre d'expériences est très important, la moyenne arithmétique aura beaucoup de chances d'être très voisine de l'espérance mathématique. Si le nombre n d'expériences n'est pas grand, on introduit une certaine erreur en remplaçant l'espérance mathématique par la moyenne arithmétique. Cette erreur est en moyenne d'autant plus importante que le nombre d'expériences est petit. Il en sera de même pour les estimations d'autres paramètres inconnus. Toutes ces estimations sont aléatoires et des erreurs sont inévitables. Il est préférable de choisir l'estimation pour laquelle ces erreurs sont minimales.

Soit une variable aléatoire X dont la loi de répartition contient un paramètre inconnu a. Il s'agit, dans le cas général, de trouver une estimation convenable pour le paramètre a d'après les résultats de n expériences indépendantes, dans chacune desquelles la variable X a pris une valeur déterminée.

Désignons les valeurs observées de X par

$$X_1, X_2, \ldots, X_n. \quad (14.1.1)$$

Celles-ci peuvent être considérées comme n échantillons de la variable aléatoire X, c'est-à-dire n variables aléatoires indépendantes, réparties chacune suivant la même loi que X.

Désignons par $\tilde{a}$ l'estimation du paramètre a. Toute estimation calculée à partir de (14.1.1) doit être une *fonction* des variables X_1, $X_2, \ldots, X_n$:

$$\tilde{a} = \tilde{a}(X_1, X_2, \ldots, X_n) \quad (14.1.2)$$

et par conséquent être elle-même une variable aléatoire. La loi de répartition de $\tilde{a}$ dépend de celle de X (et, en particulier, du paramètre inconnu a) ainsi que du nombre d'expériences n. En principe, les méthodes de la théorie des probabilités permettent de trouver cette loi de répartition.

Imposons à l'estimation $\tilde{a}$ des exigences, auxquelles elle doit satisfaire pour être considérée « bonne » en certain sens.

Il est tout naturel d'exiger que l'estimation $\tilde{a}$ tende (converge en probabilité) vers le paramètre a lorsque le nombre d'expériences augmente. Une estimation douée de cette propriété est dite *consistante*.

Ensuite, il est désirable qu'en utilisant $\tilde{a}$ au lieu de a, on ne commette pas au moins d'erreur systématique augmentant ou diminuant l'estimation obtenue c'est-à-dire que l'on ait:

$$M[\tilde{a}] = a. \quad (14.1.3)$$

Une estimation satisfaisant à cette condition est dite *non biaisée*.

Il est enfin désirable que l'estimation centrée choisie possède la variance minimale, c'est-à-dire

$$D[\tilde{a}] = \min. \tag{14.1.4}$$

Une estimation douée de cette propriété est dite *effective*.

Souvent il est difficile de satisfaire à toutes ces conditions. Par exemple, il peut se faire qu'une estimation effective existe, mais les formules de calcul sont tellement compliquées, qu'on est obligé de se contenter d'une autre estimation, dont la variance est un peu plus grande. Parfois on admet, pour simplifier les calculs, des estimations légèrement biaisées. En tout cas, le choix d'une estimation doit être précédé d'un examen critique des possibilités.

14.2. Estimations de l'espérance mathématique et de la variance

Soit une variable aléatoire X d'espérance mathématique m et de variance D inconnues. Les n expériences indépendantes en ont fourni les valeurs $X_1, X_2, \ldots, X_n$. Il y a lieu de trouver les estimations consistantes et non biaisées des paramètres m et D.

Pour l'estimation de l'espérance mathématique il est naturel de prendre la moyenne arithmétique des valeurs observées qu'auparavant nous avons désignée par m^*:

$$\tilde{m} = m^* = \frac{\sum_{i=1}^{n} X_i}{n}. \tag{14.2.1}$$

Il est facile de voir que cette estimation est consistante: en vertu de la loi des grands nombres, lorsque n augmente, $\tilde{m}$ converge en probabilité vers m. L'estimation $\tilde{m}$ est également non biaisée, en effet:

$$M[\tilde{m}] = \frac{\sum_{i=1}^{n} m}{n} = m. \tag{14.2.2}$$

La variance de cette estimation est:

$$D[\tilde{m}] = \frac{1}{n} D. \tag{14.2.3}$$

Une estimation est effective ou non suivant la loi de répartition de X. On peut montrer que si X est normalement répartie, sa variance (14.2.3) est minimale possible, et donc l'estimation $\tilde{m}$ est effective. Ceci peut ne pas être vrai pour d'autres lois de répartition.

Passons maintenant à l'estimation de la variance D. A première vue l'estimation la plus naturelle paraît être la variance statistique:

$$D^* = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^2}{n}, \tag{14.2.4}$$

où

$$\tilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}. \tag{14.2.5}$$

Nous allons vérifier que cette estimation est consistante. A cet effet exprimons-la en fonction du moment initial d'ordre deux [formule (7.4.6)] :

$$D^* = \frac{\sum_{i=1}^{n} X_i^2}{n} - \tilde{m}^2. \tag{14.2.6}$$

Le premier terme du second membre, qui est la moyenne arithmétique de n valeurs observées de la variable aléatoire X^2, converge en probabilité vers $M[X^2] = \alpha_2[X]$. Le second terme converge en probabilité vers m^2 ; donc la grandeur (14.2.6) converge en probabilité vers

$$\alpha_2[X] - m^2 = D,$$

ce qui signifie que l'estimation (14.2.4) est consistante.

Vérifions également que l'estimation D^* est non biaisée. A cet effet substituons dans la formule (14.2.6) au lieu de $\tilde{m}$ son expression (14.2.5), il vient :

$$D^* = \frac{\sum_{i=1}^{n} X_i^2}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2 =$$

$$= \frac{\sum_{i=1}^{n} X_i^2}{n} - \frac{\sum_{i=1}^{n} X_i^2}{n^2} - 2\frac{\sum_{i<j} X_i X_j}{n^2} = \frac{n-1}{n^2}\sum_{i=1}^{n} X_i^2 - 2\frac{\sum_{i<j} X_i X_j}{n^2}. \tag{14.2.7}$$

L'espérance mathématique de la grandeur (14.2.7) est :

$$M[D^*] = \frac{n-1}{n^2}\sum_{i=1}^{n} M[X_i^2] - \frac{2}{n^2}\sum_{i<j} M[X_i X_j]. \tag{14.2.8}$$

Comme la variance D^* ne dépend pas du choix de l'origine des coordonnées, plaçons celle-ci au point m. On a alors:

$$M[X_i^2]=M[\overset{\circ}{X}{}_i^2]=D;\quad \sum_{i=1}^{n} M[X_i^2]=nD, \tag{14.2.9}$$

$$M[X_iX_j]=M[\overset{\circ}{X}_i\overset{\circ}{X}_j]=K_{ij}=0. \tag{14.2.10}$$

La dernière égalité est due à ce que les expériences sont indépendantes.

En substituant (14.2.9) et (14.2.10) dans (14.2.8) on obtient:

$$M[D^*]=\frac{n-1}{n}D. \tag{14.2.11}$$

On voit que D^* est une estimation biaisée de D, car son espérance mathématique n'est pas égale à D mais lui est un peu inférieure. En utilisant l'estimation D^* au lieu de D, on commet une certaine erreur systématique en moins. Pour faire disparaître ce biais il suffit d'introduire une correction en multipliant la grandeur D^* par $\frac{n}{n-1}$. On obtient alors:

$$\frac{n}{n-1}D^*=\frac{\sum_{i=1}^{n}(X_i-\tilde{m})^2}{n}\cdot\frac{n}{n-1}=\frac{\sum_{i=1}^{n}(X_i-\tilde{m})^2}{n-1}.$$

C'est cette variance statistique « corrigée » que nous allons adopter pour l'estimation de D:

$$\tilde{D}=\frac{\sum_{i=1}^{n}(X_i-\tilde{m})^2}{n-1}. \tag{14.2.12}$$

Comme le facteur $\frac{n}{n-1}$ tend vers l'unité pour $n\to\infty$ et que l'estimation D^* est consistante, l'estimation $\tilde{D}$ le sera également *).

Souvent au lieu de (14.2.12) il est plus commode d'utiliser une autre formule qui lui est équivalente et où la variance statistique s'exprime en fonction du moment initial d'ordre deux, soit:

$$\tilde{D}=\left[\frac{\sum_{i=1}^{n}X_i^2}{n}-\tilde{m}^2\right]\frac{n}{n-1}. \tag{14.2.13}$$

Il est évident que pour des n grands les estimations biaisée D^* et non biaisée $\tilde{D}$ différeront très peu et il n'y a pas lieu d'introduire de facteur correctif.

*) L'estimation $\tilde{D}$ de la variance D n'est pas effective. Cependant dans le cas de la répartition normale elle se trouve être « asymptotiquement effective », c'est-à-dire pour $n\to\infty$ le rapport de la variance à la variance minimale possible tend vers l'unité.

Résumons les règles de traitement d'un nombre restreint de données statistiques.

Si $x_1, x_2, \ldots, x_n$ sont les valeurs que prend lors de n expériences indépendantes la variable aléatoire X d'espérance mathématique m et de variance D inconnues, pour estimer ces paramètres il y a lieu d'utiliser les valeurs approchées (estimations) suivantes:

$$\left.\begin{aligned} \tilde{m} &= \frac{\sum_{i=1}^{n} x_i}{n}; \\ \tilde{D} &= \frac{\sum_{i=1}^{n} (x_i - \tilde{m})^2}{n-1} \\ \text{ou} \quad \tilde{D} &= \left(\frac{\sum_{i=1}^{n} x_i^2}{n} - \tilde{m}^2 \right) \frac{n}{n-1}. \end{aligned}\right\} \qquad (14.2.14)$$

14.3. Intervalle de confiance. Probabilité de confiance

Dans les paragraphes précédents nous avons envisagé l'estimation du paramètre inconnu a à l'aide d'un seul nombre. Une telle estimation est dite *ponctuelle*. Dans un certain nombre de problèmes il ne suffit pas de trouver pour le paramètre une valeur numérique convenable, mais il faut estimer également sa précision et sa fiabilité. On veut savoir quelles sont les erreurs qu'entraîne le remplacement du paramètre a par son estimation ponctuelle $\tilde{a}$ et en quelle mesure on peut être sûr que ces erreurs se trouveront dans certaines limites.

Ces problèmes sont particulièrement importants lorsque le nombre de données est insuffisant. Dans ce cas l'estimation ponctuelle $\tilde{a}$ est nettement aléatoire et des erreurs importantes peuvent s'introduire lorsque l'on remplace a par $\tilde{a}$.

Pour évaluer la précision et la fiabilité d'une estimation, en mathématiques statistiques on utilise les notions d'*intervalle de confiance et de probabilité de confiance*.

Supposons que pour le paramètre a on ait expérimentalement obtenu l'estimation non biaisée $\tilde{a}$. Estimons maintenant l'erreur possible. A cet effet choisissons une probabilité assez grande β (par exemple $\beta = 0{,}9$, $0{,}95$ ou $0{,}99$) telle qu'un événement, avec la probabilité β, puisse être considéré comme certain et déterminons la valeur ε pour laquelle on a:

$$P(|\tilde{a} - a| < \varepsilon) = \beta. \qquad (14.3.1)$$

La gamme des valeurs possibles de l'erreur apparaissant lorsque l'on substitue $\tilde{a}$ à a sera alors $\pm\ \varepsilon$; des erreurs importantes en valeur absolue apparaîtront seulement avec une petite probabilité $\alpha = 1 - \beta$.

Ecrivons (14.3.1) comme suit:

$$P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \beta. \qquad (14.3.2)$$

L'égalité (14.3.2) signifie qu'avec une probabilité égale à β la valeur inconnue du paramètre a se trouvera dans l'intervalle:

$$I_\beta = (\tilde{a} - \varepsilon;\ \tilde{a} + \varepsilon). \qquad (14.3.3)$$

Soulignons une particularité de l'intervalle de confiance. Nous avons envisagé auparavant la probabilité pour une variable aléatoire de se trouver dans un intervalle déterminé. Ici la grandeur a n'est pas

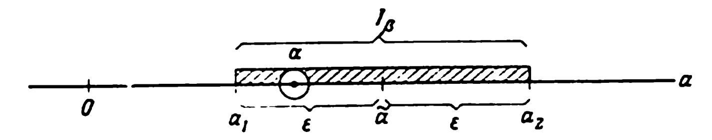


Fig. 14.3.1

aléatoire, par contre l'intervalle I_β est aléatoire. Sont aléatoires sa position sur l'axe des abscisses, déterminée par le centre $\tilde{a}$, ainsi que la longueur 2ε de cet intervalle, car la grandeur ε se calcule en général d'après des données expérimentales. C'est pourquoi, dans le cas présent, il est préférable d'interpréter la grandeur β non pas comme la probabilité pour le point a de se trouver dans l'intervalle I_β, mais comme la probabilité pour l'intervalle I_β de *recouvrir* le point a (fig. 14.3.1).

La probabilité β est généralement appelée *probabilité de confiance*, et l'intervalle I_β *intervalle de confiance* *). Les limites $a_1 = \tilde{a} - \varepsilon$ et $a_2 = \tilde{a} + \varepsilon$, de l'intervalle I_β, sont appelées *limites de confiance*.

On peut interpréter l'intervalle de confiance comme l'intervalle des valeurs du paramètre a *compatibles* avec les données expérimentales et *ne les contredisant pas*. En effet, si l'on considère un événement pratiquement impossible avec une probabilité $\alpha = 1 - \beta$, les valeurs du paramètre a pour lesquelles $|\tilde{a} - a| > \varepsilon$ sont alors en contradiction avec les données expérimentales, et celles pour lesquelles $|\tilde{a} - a| < \varepsilon$ sont compatibles avec elles.

Calculons des limites de confiance a_1 et a_2.

*) Sur la figure 14.3.1 l'intervalle de confiance est symétrique par rapport à $\tilde{a}$, comme nous allons le voir, ceci n'est pas obligatoire.

Supposons que pour le paramètre a on ait une estimation non biaisée $\tilde{a}$. Si l'on connaissait la loi de répartition de la variable $\tilde{a}$, il serait très simple de trouver l'intervalle de confiance, il suffirait de trouver une valeur ε telle que

$$P(|\tilde{a} - a| < \varepsilon) = \beta.$$

Ce qui complique le problème c'est que la loi de l'estimation $\tilde{a}$ dépend de la loi de X et par conséquent de ses paramètres inconnus (en particulier, du paramètre a lui-même).

Pour éviter cette difficulté on peut approcher dans l'expression de ε, bien que très grossièrement, les paramètres inconnus par leurs estimations ponctuelles. Lorsque le nombre d'expériences n'est pas grand (environ 20 ou 30), ce procédé donne des résultats tout à fait acceptables.

A titre d'exemple, cherchons l'intervalle de confiance de l'espérance mathématique.

Supposons que l'on ait effectué n mesures indépendantes de la variable aléatoire X, d'espérance mathématique m et de variance D inconnues. Pour ces paramètres on a obtenu les estimations suivantes:

$$\tilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}; \qquad \tilde{D} = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^2}{n-1}. \qquad (14.3.4)$$

On demande de construire l'intervalle de confiance I_β correspondant à la probabilité de confiance β pour l'espérance mathématique m de X.

Pour résoudre ce problème nous allons exploiter le fait que la grandeur $\tilde{m}$ est la somme de n variables aléatoires X_i de même loi, alors en vertu du théorème central limite, pour n suffisamment grand, sa loi de répartition est voisine de la loi normale. Dans la pratique même pour un nombre relativement petit de composantes (de l'ordre de 10 ou 20) la loi de répartition de la somme est presque normale. Supposons donc $\tilde{m}$ normalement répartie. L'espérance mathématique et la variance d'une loi normale sont respectivement m et $\frac{D}{n}$ (voir chap. 13 § 13.3). Supposons D donnée, calculons la grandeur ε_β telle que

$$P(|\tilde{m} - m| < \varepsilon_\beta) = \beta. \qquad (14.3.5)$$

En utilisant la formule (6.3.5) du chapitre 6 écrivons la probabilité dans le premier membre de (14.3.5) à l'aide de la fonction de répartition normale:

$$P(|\tilde{m} - m| < \varepsilon_\beta) = 2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right) - 1, \qquad (14.3.6)$$

où $\sigma_{\tilde{m}} = \sqrt{\frac{D}{n}}$ est l'écart quadratique moyen de l'estimation $\tilde{m}$.

L'équation

$$2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right) - 1 = \beta$$

donne la valeur ε_β:

$$\varepsilon_\beta = \sigma_{\tilde{m}} \arg \Phi^*\left(\frac{1+\beta}{2}\right), \qquad (14.3.7)$$

où $\arg \Phi^*(x)$ est la fonction inverse de $\Phi^*(x)$, c'est-à-dire une valeur de l'argument pour laquelle la fonction normale de répartition est x.

On ne connaît pas exactement la variance D entrant dans l'expression de $\sigma_{\tilde{m}}$, mais on peut l'approcher par l'estimation $\tilde{D}$ (14.3.4) et poser

$$\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}}{n}}. \qquad (14.3.8)$$

On a ainsi approximativement déterminé l'intervalle de confiance cherché:

$$I_\beta = (\tilde{m} - \varepsilon_\beta;\ \tilde{m} + \varepsilon_\beta), \qquad (14.3.9)$$

où ε_β est donné par la formule (14.3.7).

Afin d'éviter, lors du calcul de ε_β, une interpolation inverse dans les tables de la fonction $\Phi^*(x)$, il est commode de composer un tableau spécial (voir tableau 14.3.1) des

$$t_\beta = \arg \Phi^*\left(\frac{1+\beta}{2}\right)$$

pour différentes valeurs de β. La grandeur t_β donne, pour la loi normale, le nombre d'écarts quadratiques moyens qu'il y a lieu de porter à droite et à gauche du centre de dispersion pour que la probabilité de tomber dans l'intervalle obtenu soit égale à β.

L'intervalle de confiance s'exprime comme suit en fonction de t_β

$$I_\beta = (\tilde{m} - t_\beta \sigma_{\tilde{m}};\ \tilde{m} + t_\beta \sigma_{\tilde{m}}).$$

Tableau 14.3.1

β	t_β	β	t_β	β	t_β	β	t_β
0,80	1,282	0,86	1,475	0,92	1,750	0,97	2,169
0,81	1,310	0,87	1,513	0,93	1,810	0,98	2,325
0,82	1,340	0,88	1,554	0,94	1,880	0,99	2,576
0,83	1,371	0,89	1,597	0,95	1,960	0,9973	3,000
0,84	1,404	0,90	1,643	0,96	2,053	0,999	3,290
0,85	1,439	0,91	1,694				

Exemple 1. Supposons qu'on ait effectué 20 expériences sur la variable X; les résultats obtenus sont rassemblés dans le tableau 14.3.2:

Tableau 14.3.2

i	x_i	i	x_i	i	x_i	i	x_i
1	10,5	6	10,6	11	10,6	16	10,9
2	10,8	7	10,9	12	11,3	17	10,8
3	11,2	8	11,0	13	10,5	18	10,7
4	10,9	9	10,3	14	10,7	19	10,9
5	10,4	10	10,8	15	10,8	20	11,0

Trouver l'estimation $\tilde{m}$ de l'espérance mathématique $\tilde{m}$ de la variable X et l'intervalle de confiance correspondant à la probabilité de confiance $\beta = 0{,}8$.

Solution. On a:

$$\tilde{m} = \frac{1}{20} \sum_{i=1}^{20} x_i = 10{,}78.$$

Choisissons pour origine le point $x = 10$ et calculons d'après la troisième formule (14.2.14) l'estimation non biaisée $\tilde{D}$:

$$\tilde{D} = \left(\frac{13{,}38}{20} - 0{,}78^2\right)\frac{20}{19} = 0{,}064\,;$$

$$\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}}{n}} = 0{,}0564.$$

Dans le tableau 14.3.1 nous trouvons $t_\beta = 1{,}282$ et calculons:

$$\varepsilon_\beta = t_\beta \sigma_{\tilde{m}} = 0{,}072.$$

Les limites de confiance sont alors:

$$m_1 = \tilde{m} - 0{,}072 = 10{,}71\,;$$

$$m_2 = \tilde{m} + 0{,}072 = 10{,}85.$$

et l'intervalle de confiance est:

$$I_\beta = (10{,}71\,;\ 10{,}85).$$

Les valeurs du paramètre m se trouvant dans cet intervalle sont compatibles avec les données expérimentales figurant dans le tableau 14.3.2.

D'une manière analogue on peut trouver l'intervalle de confiance pour la variance.

Supposons que l'on ait effectué n expériences indépendantes sur la variable aléatoire X de paramètres inconnus m et D et que l'on ait obtenu pour la variance D l'estimation non biaisée suivante:

$$\tilde{D} = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^2}{n-1}, \qquad (14.3.11)$$

où

$$\tilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Il y a lieu de construire l'intervalle de confiance approché pour la variance.

La formule (14.3.11) montre que la grandeur $\tilde{D}$ est la somme de n variables aléatoires de la forme $\frac{(X_i - \tilde{m})^2}{n-1}$, interdépendantes, car chacune d'elles contient la variable $\tilde{m}$ dépendant de toutes les autres. Cependant on peut montrer que lorsque n augmente, la loi de répartition de la somme converge vers la loi normale. Dans la pratique pour $n = 20$ ou 30 cette loi est à peu près normale.

Supposons qu'il en soit ainsi et calculons l'espérance mathématique et la variance de cette loi. L'estimation $\tilde{D}$ étant non biaisée, on a :

$$M[\tilde{D}] = D.$$

Le calcul de la variance $D[\tilde{D}]$ est assez compliqué, c'est pourquoi nous en donnerons seulement le résultat :

$$D[\tilde{D}] = \frac{1}{n}\mu_4 - \frac{n-3}{n(n-1)}D^2, \qquad (14.3.12)$$

où μ_4 est le moment central d'ordre quatre de X.

Pour utiliser cette expression il y a lieu d'y substituer les valeurs μ_4 et D (même approchées). A la place de D on peut utiliser son estimation $\tilde{D}$. En principe le moment central d'ordre quatre peut être remplacé par son estimation, par exemple, par la grandeur

$$\mu_4^* = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^4}{n}, \qquad (14.3.13)$$

mais une telle substitution donnera une précision insuffisante, car pour un nombre d'expériences limité les moments d'ordres élevés se calculent avec des erreurs importantes. Cependant souvent dans la pratique on connaît la forme de la loi de répartition de la variable X. On peut alors essayer d'exprimer μ_4 en fonction de D.

Dans la pratique le plus souvent on rencontre la loi normale, dans ce cas le moment central d'ordre quatre s'exprime en fonction de la variance comme suit (voir chap. 6, § 6.2) :

$$\mu_4 = 3D^2,$$

et la formule (14.3.12) donne alors :

$$D[\tilde{D}] = \frac{3}{n}D^2 - \frac{n-3}{n(n-1)}D^2$$

ou

$$D[\tilde{D}] = \frac{2}{n-1} D^2. \qquad (14.3.14)$$

En rémplaçant dans (14.3.14) l'inconnue D par son estimation $\tilde{D}$ on obtient :

$$D[\tilde{D}] = \frac{2}{n-1} \tilde{D}^2, \qquad (14.3.15)$$

d'où

$$\sigma_{\tilde{D}} = \sqrt{\frac{2}{n-1}}\, \tilde{D}. \qquad (14.3.16)$$

Dans certains cas, lorsque la loi de répartition n'est pas normale, mais de forme connue, on peut également exprimer le moment μ_4 en fonction de D. Par exemple, pour la loi uniforme (voir chap. 5) on a :

$$\mu_4 = \frac{(\beta-\alpha)^4}{80}; \qquad D = \frac{(\beta-\alpha)^2}{12},$$

où (α, β) est l'intervalle sur lequel cette loi est donnée. Par conséquent :

$$\mu_4 = 1{,}8D^2.$$

La formule (14.3.12) donne :

$$D[\tilde{D}] = \frac{0{,}8n + 1{,}2}{n(n-1)} D^2,$$

d'où approximativement :

$$\sigma_{\tilde{D}} \approx \sqrt{\frac{0{,}8n+1{,}2}{n(n-1)}}\, \tilde{D}. \qquad (14.3.17)$$

Même lorsqu'on ne connaît pas la forme de la loi de X, pour estimer approximativement la grandeur $\sigma_{\tilde{D}}$ il est conseillé d'utiliser la formule (14.3.16) si seulement on n'a pas de raisons spéciales de supposer que cette loi diffère fortement de la loi normale (c'est-à-dire possède un excès positif ou négatif important).

Si d'une manière ou d'une autre on a trouvé la valeur approchée de $\sigma_{\tilde{D}}$, on peut construire l'intervalle de confiance pour la variance, tout comme dans le cas de l'espérance mathématique :

$$I_\beta = (\tilde{D} - t_\beta \sigma_{\tilde{D}};\ \tilde{D} + t_\beta \sigma_{\tilde{D}}), \qquad (14.3.18)$$

où la grandeur t_β pour différentes probabilités β est donnée dans le tableau 14.3.1.

Exemple 2. Dans les conditions de l'exemple 1 trouver approximativement l'intervalle de confiance correspondant à la probabilité de confiance $\beta = 0{,}8$ pour la variance de la variable aléatoire X, sachant que X suit une loi voisine de la loi normale.

Solution. La grandeur t_β est la même que dans l'exemple 1:

$$t_\beta = 1,282.$$

La formule (14.3.16) donne:

$$\sigma_{\tilde{D}} = \sqrt{\frac{2}{19}} \cdot 0,064 = 0,0207.$$

La formule (14.3.18) permet de trouver l'intervalle de confiance:

$$I_\beta = (0,043;\ 0,085).$$

L'intervalle correspondant des valeurs de l'écart quadratique moyen est (0,21; 0,29).

14.4. Méthodes exactes de calcul des intervalles de confiance pour les paramètres d'une variable aléatoire normale

Dans le paragraphe précédent nous avons étudié des méthodes approchées de construction des intervalles de confiance pour l'espérance mathématique et la variance. Le présent paragraphe est consacré aux méthodes précises de résolution du même problème. Soulignons que pour le calcul précis des intervalles de confiance il est indispensable de connaître à l'avance la forme de la loi de répartition de X, alors que dans le cas des méthodes approchées on peut s'en dispenser.

Les méthodes précises de calcul des intervalles de confiance reposent sur l'idée suivante. Tout intervalle de confiance peut être trouvé à partir de la condition exprimant la probabilité de réalisation de certaines inégalités contenant l'estimation $\tilde{a}$. La loi de l'estimation $\tilde{a}$ dépend dans le cas général des paramètres inconnus de la variable X. Cependant, dans certains cas on peut passer dans les inégalités de la variable aléatoire $\tilde{a}$ à une autre fonction des valeurs observées X_1, X_2, . . ., X_n dont la loi de répartition ne dépend pas des paramètres inconnus, mais seulement du nombre n d'expériences et de la forme de la loi de répartition de X. De telles fonctions aléatoires jouent un rôle important en statistique mathématique; elles sont étudiées en détail dans le cas de la loi normale.

Par exemple, on peut démontrer que si X est une variable normale, la variable aléatoire

$$T = \sqrt{n}\,\frac{\tilde{m} - m}{\sqrt{\tilde{D}}}, \tag{14.4.1}$$

où

$$\tilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad \tilde{D} = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^2}{n-1},$$

suit la *loi de répartition de Student* à $n - 1$ degrés de liberté ; la densité de probabilité de cette loi est :

$$S_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} \qquad (14.4.2)$$

où $\Gamma(x)$ est la fonction gamma

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}\,du.$$

Il a été démontré également que la variable aléatoire

$$V = \frac{(n-1)\tilde{D}}{D} \qquad (14.4.3)$$

est répartie suivant la loi du χ^2 à $n - 1$ degrés de liberté (voir chap. 7, page 137), dont la densité est donnée par la formule suivante :

$$k_{n-1}(v) = \begin{cases} \dfrac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} v^{\frac{n-1}{2}-1}e^{-\frac{v}{2}} & \text{pour } v > 0, \\ 0 & \text{pour } v < 0. \end{cases} \qquad (14.4.4)$$

Sans nous arrêter sur la démonstration des lois (14.4.2) et (14.4.4) nous allons seulement montrer comment celles-ci peuvent être utilisées pour le calcul des intervalles de confiance pour les paramètres m et D.

Supposons que n expériences indépendantes sur la variable aléatoire X, répartie suivant la loi normale de paramètres inconnus m et D, aient fourni les estimations suivantes :

$$\tilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}; \qquad \tilde{D} = \frac{\sum_{i=1}^{n}(X_i - \tilde{m})^2}{n-1}.$$

Il y a lieu de trouver pour les deux paramètres les intervalles de confiance correspondant à la probabilité de confiance β.

Nous allons tout d'abord déterminer l'intervalle de confiance pour l'espérance mathématique. Il est naturel de prendre cet intervalle symétrique par rapport à $\tilde{m}$; désignons par ε_β la moitié de cet intervalle. La grandeur ε_β doit être telle que

$$P(|\tilde{m} - m| < \varepsilon_\beta) = \beta. \qquad (14.4.5)$$

Dans le premier membre de l'égalité (14.4.5) nous allons substituer à la variable aléatoire $\tilde{m}$ la variable aléatoire T répartie suivant la

loi de Student. A cet effet nous multiplions les deux membres de l'inégalité $|\tilde{m} - m| < \varepsilon_\beta$ par le facteur positif $\frac{\sqrt{n}}{\sqrt{\tilde{D}}}$:

$$P\left(\frac{\sqrt{n}\,|\tilde{m}-m|}{\sqrt{\tilde{D}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{\tilde{D}}{n}}}\right) = \beta$$

ou, en utilisant la désignation (14.4.1):

$$P\left(|T| < \frac{\varepsilon_\beta}{\sqrt{\frac{\tilde{D}}{n}}}\right) = \beta. \qquad (14.4.6)$$

Cherchons un nombre t_β tel que

$$P(|T| < t_\beta) = \beta. \qquad (14.4.7)$$

La grandeur t_β est donnée par la condition:

$$P(|T| < t_\beta) = \int_{-t_\beta}^{t_\beta} S_{n-1}(t)\,dt = \beta. \qquad (14.4.8)$$

En vertu de la formule (14.4.2), $S_{n-1}(t)$ est une fonction paire; ainsi à partir de (14.4.8) on obtient:

$$2\int_0^{t_\beta} S_{n-1}(t)\,dt = \beta. \qquad (14.4.9)$$

L'égalité (14.4.9) donne la grandeur t_β en fonction de β. Si l'on dispose d'une table des valeurs de l'intégrale

$$\Psi(x) = 2\int_0^x S_{n-1}(t)\,dt,$$

on peut trouver la grandeur t_β par interpolation inverse dans cette table. Cependant il est plus commode de dresser à l'avance une table des valeurs de t_β (voir table 5 de l'annexe). Cette table donne les valeurs de t_β en fonction de la probabilité de confiance β et du nombre de degrés de liberté $n - 1$. Prenant t_β dans la table 5 et posant

$$\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}}{n}}, \qquad (14.4.10)$$

on trouve la moitié de la largeur de l'intervalle de confiance I_β et cet intervalle lui-même, soit:

$$I_\beta = \left(\tilde{m} - t_\beta\sqrt{\frac{\tilde{D}}{n}};\ \tilde{m} + t_\beta\sqrt{\frac{\tilde{D}}{n}}\right). \qquad (14.4.11)$$

E x e m p l e 1. Supposons qu'on ait effectué 5 expériences indépendantes sur la variable aléatoire X répartie suivant une loi normale de paramètres inconnus m et σ. Les résultats des expériences sont rassemblés dans le tableau (14.4.1).

Tableau 14.4.1

i	1	2	3	4	5
x_i	−2,5	3,4	−2,0	1,0	2,1

Trouver l'estimation $\tilde{m}$ de l'espérance mathématique et déterminer l'intervalle de confiance I_β à 90 % (c'est-à-dire l'intervalle correspondant à la probabilité de confiance $\beta = 0,9$).

S o l u t i o n. On a:

$$\tilde{m} = 0,4; \quad \tilde{D} = 6,6.$$

La table 5 de l'annexe donne pour $n - 1 = 4$, $\beta = 0,9$:

$$t_\beta = 2,13,$$

d'où:

$$\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}}{n}} \approx 2,45.$$

L'intervalle de confiance sera:

$$I_\beta = (\tilde{m} - \varepsilon_\beta;\ \tilde{m} + \varepsilon_\beta) = (-2,05;\ 2,85).$$

E x e m p l e 2. Pour les conditions de l'exemple 1 du § 14.3, supposant la variable X normalement répartie, trouver l'intervalle de confiance exact.

S o l u t i o n. Dans la table 5 de l'annexe on trouve pour $n - 1 = 19$ et $\beta = 0,8$, $t_\beta = 1,328$, d'où $\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}}{n}} \approx 0,0704$.

En comparant ce résultat avec la solution de l'exemple 1 du § 14.3 ($\varepsilon_\beta = 0,072$), on trouve que la différence est insignifiante. Les intervalles de confiance trouvés par les méthodes approchée et exacte coïncident à deux chiffres décimaux près, soit:

$$I_\beta = (10,71;\ 10,85).$$

Calculons maintenant l'intervalle de confiance pour la variance. Soit l'estimation non biaisée de la variance:

$$\tilde{D} = \frac{\sum_{i=1}^{n} (X_i - \tilde{m})^2}{n-1}.$$

Exprimons $\tilde{D}$ en fonction de la grandeur V (14.4.3) répartie suivant la loi du χ^2 (14.4.4):

$$\tilde{D} = V \frac{D}{n-1} \qquad (14.4.12)$$

Connaissant la loi de V on peut trouver l'intervalle i_β qui la contient avec une probabilité égale à β.

La figure 14.4.1 donne la loi de répartition $k_{n-1}(v)$ de V.

Le problème qui se pose est comment choisir l'intervalle i_β. Si la loi de répartition de V était symétrique (comme la loi normale ou la loi de Student), il serait naturel de prendre l'intervalle i_β symétrique par rapport à l'espérance mathématique. Dans notre cas la loi

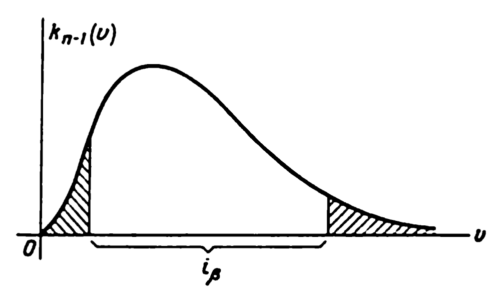


Fig. 14.4.1

$k_{n-1}(v)$, n'est pas symétrique. On s'accorde à situer l'intervalle i_β de telle sorte que les probabilités pour la variable V de se trouver à l'extérieur de l'intervalle à droite et à gauche (surfaces hachurées sur la figure 14.4.1) soient les mêmes et égales à :

$$\frac{\alpha}{2}=\frac{1-\beta}{2}.$$

Pour déterminer l'intervalle i_β répondant à cette condition nous allons utiliser la table 4 de l'annexe pour les nombres χ^2 tels que

$$P(V>\chi^2)=p,$$

V étant la variable répartie suivant la loi du χ^2 à r degrés de liberté. Ici $r=n-1$. Fixons un quelconque $r=n-1$, nous trouvons dans la ligne correspondante de la table 4 deux valeurs de χ^2 : l'une correspondant à la probabilité $p_1=\frac{\alpha}{2}$ et l'autre à la probabilité $p_2=1-\frac{\alpha}{2}$. Désignons ces valeurs par χ_1^2 et χ_2^2 qui sont respectivement les limites droite et gauche de l'intervalle i_β.

Déterminons maintenant, d'après l'intervalle i_β, l'intervalle de confiance cherché I_β pour la variance D, ayant pour limites D_1 et D_2 et recouvrant le point D avec une probabilité égale à β :

$$P(D_1<D<D_2)=\beta.$$

Proposons-nous de construire un intervalle $I_\beta=(D_1, D_2)$ tel qu'il recouvre le point D si et seulement si la variable V se trouve dans

l'intervalle i_β. Montrons que l'intervalle

$$I_\beta = \left(\frac{\tilde{D}(n-1)}{\chi_1^2}; \frac{\tilde{D}(n-1)}{\chi_2^2}\right) \qquad (14.4.13)$$

satisfait à cette condition. En effet, les inégalités:

$$\frac{\tilde{D}(n-1)}{\chi_1^2} < D; \qquad \frac{\tilde{D}(n-1)}{\chi_2^2} > D$$

sont équivalentes aux inégalités:

$$V < \chi_1^2; \qquad V > \chi_2^2,$$

celles-ci étant vérifiées avec une probabilité égale à β. Nous avons ainsi déterminé l'intervalle de confiance pour la variance, celui-ci est donné par la formule (14.4.13).

Exemple 3. Trouver l'intervalle de confiance pour les conditions de l'exemple 2 du § 14.3, sachant que la variable X est normalement répartie.

Solution. On a $\beta = 0{,}8$; $\alpha = 0{,}2$; $\frac{\alpha}{2} = 0{,}1$. Pour $r = n - 1 = 19$ dans la table 4 de l'annexe on trouve:

$$\text{pour } p_1 = \frac{\alpha}{2} = 0{,}1; \qquad \chi_1^2 = 27{,}2;$$

$$\text{pour } p_2 = 1 - \frac{\alpha}{2} = 0{,}9; \qquad \chi_2^2 = 11{,}65.$$

La formule (14.4.13) permet de déterminer l'intervalle de confiance pour la variance:

$$I_\beta = (0{,}045;\ 0{,}104).$$

L'intervalle correspondant pour l'écart quadratique moyen est (0,21 ; 0,32). Cet intervalle n'est que légèrement supérieur à l'intervalle (0,21 ; 0,29) obtenu dans l'exemple 2 du § 14.3 par la méthode approchée.

14.5. Estimation d'une probabilité d'après la fréquence de réalisation d'un événement

Dans la pratique souvent il y a lieu d'estimer la probabilité inconnue p de l'événement A d'après sa fréquence p^* dans n expériences indépendantes.

Ce problème est très voisin des problèmes étudiés dans les paragraphes précédents. En effet, la fréquence de l'événement A dans n expériences indépendantes n'est rien d'autre que la moyenne arithmétique des valeurs observées de la variable X, qui dans chaque expérience prend la valeur 1 associée à la réalisation de l'événement A, et 0 dans le cas contraire:

$$p^* = \frac{\sum_{i=1}^{n} X_i}{n}. \qquad (14.5.1)$$

Rappelons que l'espérance mathématique de la variable X est égale à p et sa variance à pq, où $q = 1 - p$. L'espérance mathématique de la moyenne arithmétique est également égale à p, soit :

$$M[p^*] = p, \tag{14.5.2}$$

c'est-à-dire que l'estimation p^* de p est non biaisée.

La variance de p^* est :

$$D[p^*] = \frac{pq}{n} \tag{14.5.3}$$

On peut montrer que cette variance est la plus petite possible, c'est-à-dire que l'estimation p^* de p est effective.

Ainsi, il semble raisonnable dans tous les cas d'adopter pour l'estimation ponctuelle de la probabilité inconnue p la fréquence p^*. Il est intéressant de connaître la précision et la fiabilité de cette estimation, c'est-à-dire de déterminer l'intervalle de confiance pour la probabilité p

Bien que ce problème soit un cas particulier du problème de la détermination de l'intervalle de confiance pour l'espérance mathématique étudié précédemment, il y a lieu de le résoudre séparément. En effet, X est une variable aléatoire discrète qui ne peut prendre que deux valeurs, 0 et 1. De plus, il existe une relation fonctionnelle entre son espérance mathématique p et la variance $pq = p(1 - p)$, ce qui facilite le calcul de l'intervalle de confiance.

Considérons d'abord le cas le plus simple, lorsque le nombre d'expériences n est relativement important et la probabilité p n'est ni trop grande ni trop petite. On peut considérer alors que la fréquence p^* de l'événement est répartie suivant une loi voisine de la loi normale *). Les calculs montrent que cette hypothèse est vraie même lorsque le nombre d'expériences n n'est pas très important : il suffit que les deux grandeurs np et nq soient supérieures à quatre. Nous supposons ces conditions remplies et donc la fréquence p^* normalement répartie. Les paramètres de cette répartition sont :

$$m_{p^*} = p; \qquad \sigma_{p^*} = \sqrt{\frac{pq}{n}}. \tag{14.5.4}$$

Supposons tout d'abord p connue. Fixons une probabilité de confiance β et déterminons l'intervalle $(p - \varepsilon_\beta,\ p + \varepsilon_\beta)$ tel que la grandeur p^* y tombe avec une probabilité β, c'est-à-dire :

$$P(|p^* - p| < \varepsilon_\beta) = \beta. \tag{14.5.5}$$

---

*) La fréquence d'un événement pour n expériences indépendantes est une variable aléatoire discrète ; lorsqu'on dit que sa loi de répartition est voisine de la loi normale, on a en vue la fonction de répartition et non pas la densité de probabilité.

La grandeur p^* étant normalement répartie, on a:

$$P(|p^*-p|<\varepsilon_\beta)=2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_p^*}\right)-1=\beta,$$

d'où, tout comme dans le § 14.3:

$$\varepsilon_\beta=\sigma_{p^*}\arg\Phi^*\left(\frac{1+\beta}{2}\right),$$

où arg Φ^* est la fonction inverse de la fonction de répartition Φ^*.

Pour calculer β introduisons la notation:

$$t_\beta=\arg\Phi^*\left(\frac{1+\beta}{2}\right).$$

On a alors

$$\varepsilon_\beta=t_\beta\sigma_{p^*}, \qquad (14.5.6)$$

où t_β est donné par le tableau 14.3.1.

Ainsi, avec une probabilité égele à β on peut affirmer que

$$|p^*-p|<t_\beta\sqrt{\frac{pq}{n}}. \qquad (14.5.7)$$

Fig. 14.5.1

En fait on ne connaît pas la grandeur p; cependant l'inégalité (14.5.7) sera vérifiée avec une probabilité égale à β que l'on connaisse ou non la probabilité p. Après avoir obtenu expérimentalement une valeur de la fréquence p^*, on peut en utilisant l'inégalité (14.5.7) trouver l'intervalle I_β recouvrant avec une probabilité égale à β le point p. En effet, on peut écrire cette inégalité de la manière suivante:

$$(p^*-p)^2<\frac{t_\beta^2}{n}p(1-p) \qquad (14.5.8)$$

et lui donner une interprétation géométrique. Portons la fréquence p^* sur l'axe des abscisses et la probabilité p sur l'axe des ordonnées (fig. 14.5.1). Le lieu géométrique des points dont les coordonnées p^* et p satisfont à l'inégalité (14.5.8) sera la partie interne de l'ellipse passant par les points (0, 0) et (1, 1) et dont les tangentes en ces points sont parallèles à l'axe Op^*. Comme la grandeur p^* ne peut être ni négative, ni supérieure à l'unité, le domaine D correspondant à l'inégalité (14.5.8) doit encore être limité à gauche et à droite par les droites $p^*=0$ et $p^*=1$. Nous pouvons maintenant pour une valeur quelconque p^*, obtenue expérimentalement, déterminer l'intervalle de confiance I_β qui avec une probabilité égale à β recouvrira la valeur inconnue p. A cet effet, menons par le point p^* une droite parallèle à l'axe des ordonnées. Les limites du domaine D

en couperont l'intervalle de confiance $I_\beta = (p_1, p_2)$. En effet, le point M d'abscisse aléatoire p^* et d'ordonnée déterministe (mais inconnue) p avec une probabilité β se trouvera à l'intérieur de l'ellipse, c'est-à-dire l'intervalle I_β recouvrira le point p avec une probabilité β.

Les dimensions et la configuration de l' « ellipse de confiance » dépendent du nombre d'expériences n. Plus n est grand, plus l'ellipse est allongée et plus l'intervalle de confiance est étroit.

On peut trouver les limites de confiance p_1 et p_2 à partir de la relation (14.5.8) en y remplaçant l'inégalité par l'égalité. En résolvant l'équation du second degré en p on trouve deux racines:

$$\left.\begin{aligned} p_1 &= \frac{p^* + \frac{1}{2}\frac{t_\beta^2}{n} - t_\beta\sqrt{\frac{p^*(1-p^*)}{n} + \frac{1}{4}\frac{t_\beta^2}{n^2}}}{1+\frac{t_\beta^2}{n}}; \\ p_2 &= \frac{p^* + \frac{1}{2}\frac{t_\beta^2}{n} + t_\beta\sqrt{\frac{p^*(1-p^*)}{n} + \frac{1}{4}\frac{t_\beta^2}{n^2}}}{1+\frac{t_\beta^2}{n}}. \end{aligned}\right\} \qquad (14.5.9)$$

L'intervalle de confiance pour la probabilité p sera

$$I_\beta = (p_1, p_2).$$

Exemple 1. La fréquence p^* de l'événement A dans une série de 100 expériences s'est trouvée égale à 0,78. Déterminer l'intervalle de confiance à 90 % pour la probabilité de l'événement A.

Solution. Vérifions tout d'abord s'il est possible d'utiliser la loi normale; à cet effet calculons les grandeurs np et nq. En supposant $p \approx p^*$ on obtient:

$$np \approx np^* = 78; \quad nq \approx n(1 - p^*) = 22.$$

Ces deux grandeurs sont nettement supérieures à quatre, donc la loi normale est applicable. Le tableau 14.3.1 donne, pour $\beta = 0{,}9$, $t_\beta = 1{,}643$. On obtient alors à partir des formules (14.5.9):

$$p_1 = 0{,}705; \quad p_2 = 0{,}840; \quad I_\beta = (0{,}705;\ 0{,}840).$$

Notons que lorsque n augmente, les grandeurs $\frac{t_\beta^2}{n}$ et $\frac{1}{4}\frac{t_\beta^2}{n}$ dans les formules (14.5.9) tendent vers zéro; à la limite ces formules deviennent:

$$\left.\begin{aligned} p_1 &= p^* - t_\beta\sqrt{\frac{p^*(1-p^*)}{n}}, \\ p_2 &= p^* + t_\beta\sqrt{\frac{p^*(1-p^*)}{n}}. \end{aligned}\right\} \qquad (14.5.10)$$

Ces formules peuvent être obtenues directement, en utilisant la méthode approchée du § 14.3 de détermination de l'intervalle de

confiance pour l'espérance mathématique. Les formules (14.5.10) peuvent être utilisées pour des n grands (de l'ordre d'une centaine), si seulement la probabilité p n'est ni trop grande et ni trop petite (lorsque les deux grandeurs np et nq sont de l'ordre de 10 au plus)

Exemple 2. Supposons que l'on ait effectué 200 expériences; la fréquence de l'événement A est $p^* = 0{,}34$. Déterminer approximativement l'intervalle de confiance à 85 % à l'aide des formules (14.5.10). Comparer le résultat obtenu avec celui des formules exactes (14.5.9).

Solution. $\beta = 0{,}85$; le tableau 14.3.1 donne $t_\beta = 1{,}439$. En introduisant le facteur

$$\sqrt{\frac{p^*(1-p^*)}{n}} \approx 0{,}0335$$

on obtient :

$$t_\beta \sqrt{\frac{p^*(1-p^*)}{n}} \approx 0{,}048,$$

d'où la valeur approchée de l'intervalle de confiance:

$$I_\beta \approx (0{,}292;\ 0{,}388).$$

Les formules (14.5.9) permettent de trouver des valeurs plus exactes $p_1 = 0{,}294$; $p_2 = 0{,}389$ qui diffèrent peu des valeurs approchées.

Ci-dessus nous avons déterminé l'intervalle de confiance pour un nombre suffisamment grand d'expériences, lorsque la fréquence pouvait être considérée normalement répartie. Si le nombre d'expériences est petit (ou si la probabilité p est très grande ou très petite) cette approximation n'est pas vraie. Dans ce cas l'intervalle de confiance doit être déterminé non pas à partir de la loi approchée, mais de la loi exacte de répartition de la fréquence. Comme on peut le voir la fréquence suit une loi binomiale étudiée dans les chapitres 3 et 4. En effet, le nombre d'apparitions de l'événement A dans n expériences est réparti suivant la loi binomiale: la probabilité pour l'événement A d'apparaître exactement m fois est égale à :

$$P_{m,n} = C_n^m p^m q^{n-m}, \tag{14.5.11}$$

et la fréquence p^* n'est rien d'autre que le nombre d'apparitions de l'événement divisé par le nombre d'expériences.

A partir de cette répartition on peut déterminer l'intervalle de confiance I_β tout comme dans le cas de la loi normale pour des n grands (page 317).

Supposons tout d'abord la probabilité p connue et calculons l'intervalle de fréquences p_1^*, p_2^*, dans lequel avec une probabilité $\beta = 1 - \alpha$ se trouve la fréquence p^* de l'événement.

Pour un n grand nous avons utilisé la loi normale et nous avons pris l'intervalle de confiance symétrique par rapport à l'espérance mathématique. La répartition binomiale (14.5.11) n'est pas symétrique. De plus il peut se faire que l'intervalle où l'événement tombe avec une probabilité égale exactement à β peut ne pas exister (car la

fréquence est une variable aléatoire discrète). C'est pourquoi nous allons prendre pour intervalle p_1^*, p_2^* le plus petit des intervalles, tel que la probabilité pour l'événement de se trouver à gauche et à droite soit supérieure à $\frac{\alpha}{2}$.

Tout comme nous avons déterminé le domaine D pour la loi normale (fig. 14.5.1), nous pouvons pour chaque n et un β donné déterminer le domaine à l'intérieur duquel la valeur de la probabilité p est compatible avec la valeur observée de la fréquence p^*.

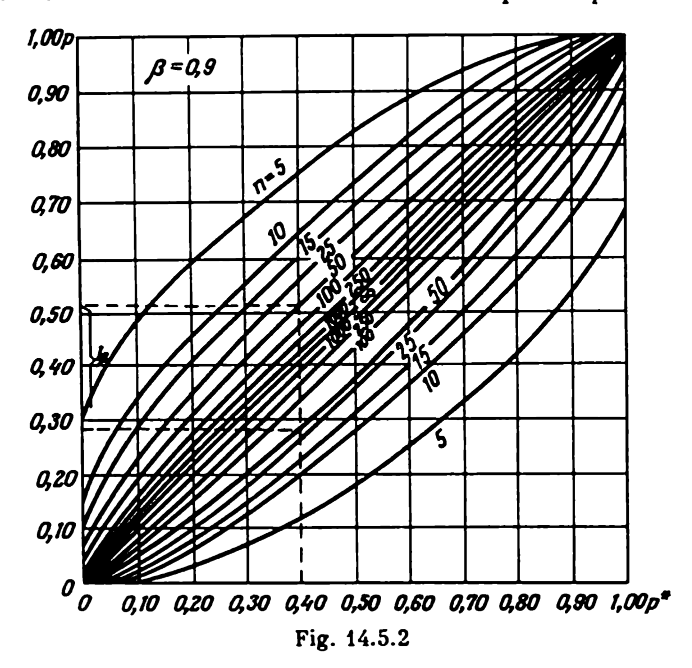


Fig. 14.5.2

Sur la figure 14.5.2 on peut voir les courbes limitant ces domaines pour différents n pour une probabilité de confiance $\beta = 0,9$. Sur l'axe des abscisses on a porté la fréquence p^*, sur l'axe des ordonnées, la probabilité p. Chaque paire de courbes correspondant à n donné détermine l'intervalle de confiance associé à une valeur donnée de la fréquence. En toute rigueur les limites des domaines doivent être échelonnées vu que la fréquence est discrète, mais pour plus de commodité elles sont représentées comme des courbes continues.

Pour déterminer l'intervalle de confiance I_β à l'aide de ces courbes il y a lieu de procéder comme suit : sur l'axe des abscisses on porte la valeur observée de la fréquence p^*, on mène par ce point une droite parallèlement à l'axe des ordonnées et on marque les points d'intersection de cette droite avec le couple de courbes correspondant à un

nombre d'expériences n; les projections de ces points sur l'axe des ordonnées seront les limites p_1, p_2 de l'intervalle de confiance I_β.

Pour un n fixé les courbes limitant le « domaine de confiance » sont données par les équations suivantes:

$$\sum_{m=k}^{n} C_n^m p^m (1-p)^{n-m} = \frac{\alpha}{2}; \qquad (14.5.12)$$

$$\sum_{m=0}^{k} C_n^m p^m (1-p)^{n-m} = \frac{\alpha}{2}, \qquad (14.5.13)$$

où

$$k = np^*$$

est le nombre de réalisations de l'événement.

La résolution de l'équation (14.5.12) par rapport à p donne la limite inférieure p_1 « du domaine de confiance »; d'une manière analogue on peut trouver p_2 à partir de (14.5.13).

Pour éviter de résoudre ces équations chaque fois qu'il est besoin, on peut tabuler (ou représenter graphiquement) les solutions pour plusieurs valeurs fréquentes de la probabilité de confiance β. Par exemple, dans l'ouvrage de I. Dounine-Barkovski « Théorie des probabilités et statistique mathématique. Applications techniques » (en russe) on peut trouver des tables donnant p_1 et p_2 pour $\beta = 0,95$ et $\beta = 0,99$; le graphique de la figure 14.5.2 est également emprunté à cet ouvrage.

Exemple 3. Avec une probabilité de confiance $\beta = 0,9$ trouver les limites de confiance p_1 et p_2 pour la probabilité d'un événement si lors de n expériences sa fréquence s'est trouvée être égale à $p^* = 0,4$.

Solution. Par construction (pointillé sur la figure 14.5.2) pour $p^* = 0.4$ et $n = 50$ on trouve $p_1 \approx 0,28$; $p_2 \approx 0,52$.

La méthode des intervalles de confiance permet également de répondre, ne serait-ce que d'une manière approchée, à une autre question très importante pour la pratique: quel est le nombre d'expériences n pour lequel, avec une probabilité de confiance β, l'erreur qu'on fait en remplaçant la probabilité par la fréquence ne soit pas supérieure à une certaine valeur ?

Pour résoudre des problèmes de ce genre il est peu commode d'utiliser directement des graphiques du type de ceux de la figure 14.5.2, il y a lieu de les reconstruire, en représentant les limites de confiance en fonction du nombre d'expériences n.

Exemple 4. Lors de 25 expériences l'événement A s'est produit 12 fois. Trouver approximativement le nombre d'expériences nécessaires pour qu'avec une probabilité $\beta = 0,9$ l'erreur commise en remplaçant la probabilité par la fréquence ne soit pas supérieure à 20 %.

S o l u t i o n. L'erreur maximale admissible est:

$$\Delta = 0{,}2 \cdot 0{,}48 = 0{,}096 \approx 0{,}1.$$

A partir des courbes de la figure 14.5.2 construisons un nouveau graphique en portant sur l'axe des abscisses le nombre d'expériences n et sur l'axe des ordonnées les limites de confiance pour la probabilité (fig. 14.5.3). La médiane parallèle à l'axe des abscisses correspond à la fréquence observée de l'événement $p^* = \frac{12}{25} = 0{,}48$. Au-dessus et au-dessous de la droite $p = p^* = 0{,}48$ sont tracées les courbes $p_1(n)$ et $p_2(n)$ représentant les limites de confiance supérieure et inférieure en fonction de n. Le domaine entre les courbes déterminant l'intervalle de confiance est hachuré. De part et d'autre de la droite $p = 0{,}48$ les hachures croisées délimitent le domaine plus étroit correspondant à 20 % de l'erreur admissible. La figure 14.5.3 montre que l'erreur entre dans les limites imposées pour un nombre d'expériences de l'ordre de 100.

Fig. 14.5.3

Notons qu'après un nombre requis d'expériences il peut s'avérer nécessaire de vérifier de nouveau la précision de calcul de la probabilité d'après la fréquence, car en général on obtient une autre valeur de la fréquence p^* différente de celle observée dans les expériences antérieures. Il peut se faire que le nombre d'expériences soit encore insuffisant pour assurer la précision requise, et qu'il y aura lieu de l'augmenter un peu. Cependant cette première approximation peut être mise à la base des calculs préalables d'une série d'expériences du point de vue du temps nécessaire, du coût, etc.

Dans la pratique parfois il y a lieu de déterminer l'intervalle de confiance pour la probabilité d'un événement, lorsque la fréquence obtenue expérimentalement est égale à zéro. Généralement c'est le cas des expériences où la probabilité de l'événement étudié est très petite (ou au contraire très grande, ce qui rend très petite la probabilité de l'événement contraire).

Supposons, par exemple, qu'on effectue des essais de défaillance d'un appareil. Lors des essais l'appareil n'a eu aucune défaillance. Trouver la probabilité maximale possible de défaillance de l'appareil étudié.

Posons ce problème sous une forme générale. Aucune des n expériences indépendantes n'a réalisé l'événement A. Pour une probabilité de confiance β donnée, déterminer l'intervalle de confiance pour la probabilité p de l'événement A, plus exactement trouver sa limite

supérieure p_2, la limite inférieure p_1 étant évidemment égale à zéro.

Le problème posé est un cas particulier du problème général de la détermination de l'intervalle de confiance d'une probabilité, mais vu ses particularités, il mérite une étude spéciale. Tout d'abord la méthode approchée de détermination de l'intervalle de confiance (dans laquelle on suppose normale la loi de répartition de la fréquence), exposée au début de ce paragraphe n'est pas applicable ici, car la probabilité p est très petite. La méthode exacte basée sur la loi binomiale reste applicable, mais admet d'importantes simplifications.

Nous allons raisonner de la manière suivante. Lors de n expériences a été réalisé l'événement B consistant dans la non-apparition de l'événement A. Il y a lieu de trouver la valeur maximale $p = p_2$ « compatible » avec l'événement observé B, supposant « incompatibles » avec B les valeurs de p pour lesquelles la probabilité de l'événement B est inférieure à $\alpha = 1 - \beta$.

Il est évident que quelle que soit toute la probabilité p de l'événement A, la probabilité de l'événement observé B est égale à :

$$P(B) = (1 - p)^n.$$

En posant $P(B) = \alpha$ on obtient l'équation suivante pour p_2:

$$(1 - p_2)^n = 1 - \beta, \qquad (14.5.14)$$

d'où

$$p_2 = 1 - \sqrt[n]{1 - \beta}. \qquad (14.5.15)$$

E x e m p l e 5. La probabilité d'une explosion spontanée des vapeurs d'essence secouée est inconnue mais peut être supposée assez petite. On a effectué 100 expériences, aucune explosion n'a été enregistrée. Déterminer la limite supérieure de l'intervalle de confiance à 90 % pour la probabilité d'explosion.

S o l u t i o n. La formule (14.5.15) donne:

$$p_2 = 1 - \sqrt[100]{1 - 0{,}9} = 1 - \sqrt[100]{0{,}1},$$

$$\log \sqrt[100]{0{,}1} = 0{,}01 \log 0{,}1 = \bar{1}{,}9900,$$

$$\sqrt[100]{0{,}1} \approx 0{,}977, \quad p_2 = 1 - 0{,}977 = 0{,}023.$$

Considérons un autre problème lié au précédent. L'événement A de petite probabilité p n'a pas été observé lors de n expériences. Pour une probabilité de confiance β donnée, trouver le nombre d'expériences n à effectuer pour que la limite de confiance supérieure de la probabilité de l'événement soit égale à une valeur donnée p_2.

La solution de ce problème est donnée directement par la formule (14.5.14) :

$$n = \frac{\log (1 - \beta)}{\log (1 - p_2)}. \qquad (14.5.16)$$

E x e m p l e 6. Dans combien d'expériences il faut constater un travail sans défaillance d'un appareil pour affirmer avec une certitude de 95 % que, dans la pratique, une panne surviendra dans moins de 5 % des cas?

S o l u t i o n. D'après la formule (14.5.16) pour $\beta = 0,95$, $p_2 = 0,05$ on a:

$$n = \frac{\log 0,05}{\log 0,95} \approx 58,4.$$

En arrondissant par excès on obtient:

$$n = 59.$$

Vu le caractère approximatif des calculs effectués, on peut utiliser, à la place des formules (14.5.15) et (14.5.16), des formules approchées plus simples. Ces dernières peuvent être obtenues en supposant que le nombre d'apparitions de l'événement A, lors de n expériences, est réparti suivant une loi de Poisson d'espérance mathématique $a = np$. Cette hypothèse est approximativement vérifiée lorsque la probabilité p est relativement petite (voir chap. 5, § 5.9). On a:

$$P(B) \approx e^{-np},$$

alors au lieu de la formule (14.5.15) on obtient:

$$p_2 \approx \frac{-\ln(1-\beta)}{n} \tag{14.5.17}$$

et au lieu de la formule (14.5.16):

$$n \approx -\frac{\ln(1-\beta)}{p_2}. \tag{14.5.18}$$

E x e m p l e 7. Trouver la valeur approximative de p_2 pour les conditions de l'exemple 5.

S o l u t i o n. D'après la formule (14.5.14):

$$p_2 = \frac{-\ln 0,1}{100} = \frac{2,303}{100} = 0,023$$

c'est-à-dire on retrouve le résultat obtenu à l'aide de la formule exacte de l'exemple 5.

E x e m p l e 8. Trouver la valeur approximative de n pour les conditions de l'exemple 6.

S o l u t i o n. La formule (14.5.18) donne:

$$n \approx \frac{-\ln 0,05}{0,05} = \frac{2,996}{0,05} = 59,9.$$

En arrondissant par excès on trouve $n = 60$, ce qui diffère peu du résultat $n = 59$ obtenu dans l'exemple 6.

14.6. Estimations des caractéristiques numériques d'un système de variables aléatoires

Dans les §§ 14.1 à 14.4 nous avons étudié des problèmes d'estimation des caractéristiques numériques d'une seule variable aléatoire pour un nombre d'expériences limité et de détermination des intervalles de confiance pour ces caractéristiques.

Des problèmes analogues apparaissent lors du traitement d'un nombre limité de données relatives à deux ou plusieurs variables aléatoires.

Ici nous allons nous limiter à l'étude des estimations ponctuelles des caractéristiques d'un système.

Soit tout d'abord le cas de deux variables aléatoires.

Les n expériences indépendantes effectuées sur le système de variables aléatoires (X, Y) ont fourni les résultats suivants:

$$(x_1, y_1); (x_2, y_2); \ldots ; (x_n, y_n).$$

Il y a lieu de trouver les estimations des caractéristiques numériques du système, à savoir des espérances mathématiques m_x, m_y, des variances D_x, D_y et de la covariance K_{xy}.

Ce problème se résout comme dans le cas d'une seule variable aléatoire. Les moyennes arithmétiques seront les estimations non biaisées des espérances mathématiques:

$$\tilde{m}_x = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \tilde{m}_y = \frac{\sum_{i=1}^{n} y_i}{n}, \tag{14.6.1}$$

tandis que pour les éléments de la matrice de corrélation on a:

$$\left.\begin{aligned} \tilde{D}_x &= \frac{\sum_{i=1}^{n} (x_i - \tilde{m}_x)^2}{n-1}; \\ \tilde{D}_y &= \frac{\sum_{i=1}^{n} (y_i - \tilde{m}_y)^2}{n-1}; \\ \tilde{K}_{xy} &= \frac{\sum_{i=1}^{n} (x_i - \tilde{m}_x)(y_i - \tilde{m}_y)}{n-1}. \end{aligned}\right\} \tag{14.6.2}$$

La démonstration est analogue à celle donnée dans le § 14.2.

Lors du calcul direct des estimations des variances et de la covariance il est souvent commode d'utiliser la relation existant entre les moments statistiques centrés et initiaux, à savoir:

$$\left.\begin{aligned} D_x^* &= \alpha_2^* [X] - (m_x^*)^2; \\ D_y^* &= \alpha_2^* [Y] - (m_y^*)^2; \\ K_{xy}^* &= \alpha_{1,1}^* [X, Y] - m_x^* m_y^*, \end{aligned}\right\} \tag{14.6.3}$$

où

$$\left.\begin{aligned} m_x^* = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \alpha_2^*[X] = \frac{\sum_{i=1}^{n} x_i^2}{n}; \\ m_y^* = \frac{\sum_{i=1}^{n} y_i}{n}; \quad \alpha_2^*[Y] = \frac{\sum_{i=1}^{n} y_i^2}{n}; \\ \alpha_{1,1}^*[X, Y] = \frac{\sum_{i=1}^{n} x_i y_i}{n}. \end{aligned}\right\} \qquad (14.6.4)$$

Ayant calculé les moments statistiques à l'aide des formules (14.6.3) on peut trouver les estimations non biaisées des éléments de la matrice de corrélation à l'aide des formules suivantes :

$$\left.\begin{aligned} \widetilde{D}_x = D_x^* \frac{n}{n-1}; \\ \widetilde{D}_y = D_y^* \frac{n}{n-1}; \\ \widetilde{K}_{xy} = K_{xy}^* \frac{n}{n-1}. \end{aligned}\right\} \qquad (14.6.5)$$

Exemple. Soit un système de deux variables aléatoires : X, la taille d'un homme, en cm, et Y, sa pointure. On a effectué 20 expériences. Chaque i-ème expérience a fourni un couple de valeurs (x_i, y_i). Les résultats sont rassemblés dans le tableau 14.6.1.

Tableau 14.6.1

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	172	41	6	170	40	11	170	42	16	187	44
2	174	43	7	174	42	12	185	45	17	166	40
3	175	41	8	175	43	13	180	42	18	167	41
4	181	44	9	164	40	14	172	39	19	171	42
5	168	38	10	168	41	15	174	43	20	179	43

Trouver les estimations des caractéristiques numériques du système (X, Y), à savoir des espérances mathématiques, des variances, de la covariance et du coefficient de corrélation.

Solution. Pour simplifier les calculs, choisissons l'origine des coordonnées au point $x_0 = 170$, $y_0 = 40$. Ceci nous oblige de passer des variables X et Y aux variables $X' = X - 170$, $Y' = Y - 40$ (voir tableau 14.6.2). Seules les espérances mathématiques changent avec ce changement d'origine, les variances et la covariance restent inchangées.

Tableau 14.6.2

i	x'_i	y'_i	i	x'_i	y_i	i	x'_i	y'_i	i	x'_i	y'_i
1	2	1	6	0	0	11	0	2	16	17	4
2	4	3	7	4	2	12	15	5	17	−4	0
3	5	1	8	5	3	13	10	2	18	−3	1
4	1	4	9	−6	0	14	2	−1	19	1	2
5	−2	−2	10	−2	1	15	4	3	20	9	3

Portons les valeurs obtenues sur un graphique (fig. 14.6.1). La disposition des points sur la figure 14.6.1 laisse entrevoir une certaine relation (une corrélation positive) entre les variables du système. Estimons les caractéristiques numériques du système.

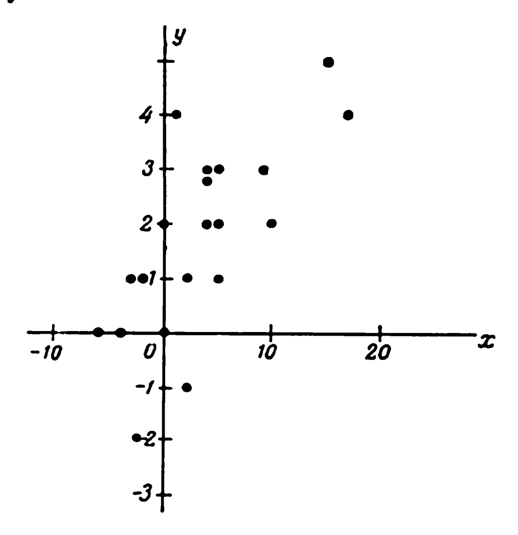


Fig. 14.6.1

Calculons à l'aide des formules (14.6.1) les valeurs moyennes de X' et Y' (les estimations des espérances mathématiques):

$$\tilde{m}_{x'} = \frac{1}{20} \sum_{i=1}^{20} x'_i = 3,1,$$

$$\tilde{m}_{y'} = \frac{1}{20} \sum_{i=1}^{20} y'_i = 1,7,$$

et, en reportant l'origine à zéro:

$$\tilde{m}_{x'} = 170 + 3,1 = 173,1,$$

$$\tilde{m}_{y'} = 40 + 1,7 = 41,7.$$

Calculons les moments statistiques initiaux deux des variables X' et Y':

$$\alpha_2^*[X'] = \frac{1}{20}\sum_{i=1}^{20} x_i'^2 = 43.6,$$

$$\alpha_2^*[Y'] = \frac{1}{20}\sum_{i=1}^{20} y_i'^2 = 5,9.$$

Trouvons d'après les formules (14.6.3) les variances statistiques des variables X' et Y' (égales à celles du X et Y):

$$D_x^* = 43,6 - 3,1^2 = 33,99,$$

$$D_y^* = 5,9 - 1,7^2 = 3,01.$$

Pour trouver les estimations non biaisées il y a lieu de multiplier les variances statistiques par $\frac{n}{n-1} = \frac{20}{19}$. Il vient:

$$\tilde{D}_x \approx 23,38; \quad \tilde{D}_y = 3,17.$$

Les estimations des écarts quadratiques moyens ont respectivement pour valeurs:

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x} \approx 4,83, \quad \sigma_y = \sqrt{\tilde{D}_y} = 1,78.$$

La dernière des formules (14.6.4) permet de trouver le moment statistique initial d'ordre 1,1:

$$\alpha_{1,1}^*[X', Y'] = \frac{1}{20}\sum_{i=1}^{20} x_i' y_i' = 12,35$$

et la covariance statistique des variables X', Y' (égale à celle des X, Y):

$$K_{xy}^* = \alpha_{1,1}^*[X', Y'] - m_{x'} m_{y'} = 12,35 - 3,1 \cdot 1,7 = 7,08.$$

L'estimation non biaisée s'obtient en multipliant par $\frac{20}{19}$:

$$\tilde{K}_{xy} \approx 7,44.$$

Ensuite calculons l'estimation du coefficient de corrélation:

$$\tilde{r}_{xy} = \frac{\tilde{K}_{xy}}{\tilde{\sigma}_x \tilde{\sigma}_y} = \frac{7,44}{4,83 \cdot 1,78} \approx 0,865.$$

La valeur assez grande de $\tilde{r}_{xy}$ témoigne de l'existence d'une relation linéaire assez importante entre les variables aléatoires X et Y.

Nous allons passer au cas d'un nombre arbitraire de variables aléatoires.

Soit un système de m variables aléatoires:

$$(X_1, X_2, \ldots, X_m).$$

Le système a été soumis à n expériences indépendantes; les résultats sont rassemblés dans un tableau, dont chaque ligne contient m valeurs prises par les variables aléatoires $X_1, X_2, \ldots, X_m$ dans une expérience (tableau 14.6.3).

Tableau 14.6.3

i	x_1	x_2	...	x_k	...	x_m
1	x_{11}	x_{21}	...	x_{k1}	...	x_{m1}
2	x_{12}	x_{22}	...	x_{k2}	...	x_{m2}
.	.	.		.		.
.	.	.	...	.	...	.
.	.	.		.		.
i	x_{1i}	x_{2i}	...	x_{ki}	...	x_{mi}
.	.	.		.		.
.	.	.	...	.	...	.
.	.	.		.		.
n	x_{1n}	x_{2n}	...	x_{kn}	...	x_{mn}

Le double indice dont est affectée chaque valeur figurant au tableau se déchiffre comme suit : le premier chiffre désigne le numéro de la variable aléatoire, le second le numéro de l'expérience, ainsi x_{ki} est la valeur prise par la variable X_k dans l' i-ème expérience.

Il y a lieu de trouver les estimations des caractéristiques numériques du système, à savoir des espérances mathématiques $m_{x_1}, m_{x_2}, \ldots$ $\ldots, m_{x_m}$ et des éléments de la matrice de corrélation :

$$\|K_{ij}\| = \left\| \begin{matrix} K_{11} K_{12} \ldots K_{1m} \\ K_{22} \ldots K_{2m} \\ \ldots\ldots\ldots\ldots \\ K_{mm} \end{matrix} \right\|.$$

Sur la diagonale principale de la matrice de corrélation on a évidemment les variances des variables aléatoires $X_1, X_2, \ldots, X_m$:

$$K_{11} = D_1; \; K_{22} = D_2; \ldots; \; K_{mm} = D_m.$$

Les estimations des espérances mathématiques sont données par les moyennes arithmétiques :

$$\tilde{m}_{x_k} = \frac{\sum_{i=1}^{n} x_{ki}}{n} \qquad (k = 1, 2, \ldots, m). \tag{14.6.6}$$

Les estimations non biaisées des variances se calculent à l'aide des formules :

$$\tilde{D}_k = \frac{\sum_{i=1}^{n} (x_{ki} - \tilde{m}_{x_k})^2}{n-1} \tag{14.6.7}$$

et celles des covariances, à l'aide des formules:

$$\widetilde{K}_{kl} = \frac{\sum_{i=1}^{n} (x_{ki} - \widetilde{m}_{x_k})(x_{li} - \widetilde{m}_{x_l})}{n-1}. \tag{14.6.8}$$

A partir de ces données on peut trouver les estimations des éléments de la matrice de corrélation normée:

$$\widetilde{r}_{kl} = \frac{\widetilde{K}_{kl}}{\widetilde{\sigma}_k \widetilde{\sigma}_l}, \tag{14.6.9}$$

où

$$\widetilde{\sigma}_k = \sqrt{\widetilde{D}_k}; \quad \widetilde{\sigma}_l = \sqrt{\widetilde{D}_l}. \tag{14.6.10}$$

Souvent il est plus commode de calculer les estimations des variances et des covariances non pas d'après les formules (14.6.7), (14.6.8) mais à l'aide des moments initiaux, soit:

$$\widetilde{D}_x = \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} x_{ki}^2 - \widetilde{m}_{x_k}^2 \right. ; \tag{14.6.11}$$

$$\widetilde{K}_{kl} = \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} x_{ki} x_{li} - \widetilde{m}_{x_k} \widetilde{m}_{x_l} \right. . \tag{14.6.12}$$

E x e m p l e. On a effectué 10 observations sur un système de 5 variables aléatoires

$$X_1, X_2, X_3, X_4, X_5.$$

Les résultats des observations sont réunis dans le tableau 14.6.4 (i — numéro de l'observation, k — numéro de la variable aléatoire).

Tableau 14.6.4

i \ k	1	2	3	4	5
1	−20	−15	−8	−6	−2
2	40	60	120	125	130
3	−25	−30	−20	−10	2
4	−100	−75	−35	2	2
5	−40	−30	−25	−30	−45
6	80	30	25	10	2
7	14	25	25	30	10
8	80	75	60	10	−4
9	−70	−60	−30	−10	0
10	2	4	10	12	4

Trouver les valeurs convenables (estimations) des caractéristiques numériques du système (X_1, X_2, X_3, X_4, X_5), à savoir des espérances mathématiques et des éléments des matrices de corrélation simple et normée.

S o l u t i o n. Les estimations des espérances mathématiques des variables aléatoires X_1, X_2, X_3, X_4, X_5 sont données par les moyennes arithmétiques calculées par colonne, soit:

$$\tilde{m}_1 = -3,9;\quad \tilde{m}_2 = -1,6;\quad \tilde{m}_3 = 12,2;\quad \tilde{m}_4 = 13,3;\quad \tilde{m}_5 = 9,9.$$

Les estimations des variances s'obtiennent à l'aide de la formule (14.6.11) en élevant au carré tous les éléments d'une colonne du tableau 14.6.4, en prenant la somme et la moyenne de ces carrés et en retranchant du résultat le carré de l'espérance mathématique correspondante, puis en multipliant par le facteur correcteur $\frac{n}{n-1} = \frac{10}{9}$. On obtient ainsi:

$$\tilde{D}_1 = 35,5\cdot 10^2;\ \tilde{D}_2 = 24,4\cdot 10^2;\ \tilde{D}_3 = 23,4\cdot 10^2;\ \tilde{D}_4 = 19,6\cdot 10^2;\ \tilde{D}_5 = 20,1\cdot 10^2,$$

d'où

$$\tilde{\sigma}_1 = 60;\quad \tilde{\sigma}_2 = 49;\quad \tilde{\sigma}_3 = 48;\quad \tilde{\sigma}_4 = 44;\quad \tilde{\sigma}_5 = 45.$$

Pour obtenir les estimations $\tilde{K}_{kl}$ des covariances il y a lieu d'utiliser la formule (14.6.12). Par exemple, pour trouver $\tilde{K}_{13}$ (estimation de la covariance des variables aléatoires X_1 et X_3) il faut multiplier les éléments d'une même ligne de la première colonne par ceux de la troisième du tableau 14.6.4, sommer tous ces produits et diviser la somme par $n = 10$, retrancher du résultat obtenu le produit des estimations correspondantes des espérances mathématiques $\tilde{m}_1\cdot\tilde{m}_3$ et multiplier la différence obtenue par le facteur correcteur $\frac{n}{n-1} = \frac{10}{9}$. Nous calculons ainsi les éléments de la matrice de corrélation. Vu sa symétrie nous n'écrivons que la moitié de la matrice, ayant sorti le facteur commun 10^2:

$$\tilde{K}_k = 10^2\cdot \begin{Vmatrix} 35,5 & 27,8 & 22,9 & 11,0 & 8,3 \\ & 24,4 & 21,4 & 13,0 & 10,2 \\ & & 23,4 & 17,4 & 17,2. \\ & & & 19,6 & 19,3 \\ & & & & 20,1 \end{Vmatrix}$$

En divisant chaque élément de cette matrice par le produit des estimations des écarts quadratiques moyens correspondants $\tilde{\sigma}_k\tilde{\sigma}_l$ nous obtenons la matrice de corrélation normée, soit:

$$r_{kl} = \begin{Vmatrix} 1,00 & 0,94 & 0,76 & 0,42 & 0,31 \\ & 1,00 & 0,89 & 0,59 & 0,46 \\ & & 1,00 & 0,81 & 0,80 \\ & & & 1,00 & 0,97 \\ & & & & 1,00 \end{Vmatrix}$$

14.7. Traitement des données des observations sur un couple de variables aléatoires normales

Dans la pratique il y a souvent lieu de traiter les résultats des observations sur deux variables aléatoires normalement réparties. Supposons qu'à l'aide d'un certain appareil on détermine n fois les coordonnées d'un certain objectif dans le plan. Diverses causes aléatoires font que les résultats des observations ne se répètent pas mais changent d'une expérience à l'autre (d'une observation à l'autre).

Après n observations on obtient n points dans le plan xOy (fig. 14.7.1).

Si les erreurs d'observation se composent d'un grand nombre d'erreurs « élémentaires » indépendantes ou faiblement liées, dues chacune à une cause particulière, on peut considérer les valeurs mesurées des coordonnées (X, Y) comme normalement réparties. Le problème se pose d'estimer, sur la base de l'expérience, les paramètres de cette loi, à savoir les coordonnées du centre de dispersion m_x, m_y, la direction des axes principaux de dispersion $O\xi$, $O\eta$, les écarts quadratiques moyens principaux σ_ξ, σ_η. Le problème est analogue dans le cas du traitement des observations des résultats des tirs par obus: les points d'impact ont une certaine dispersion dont il y a lieu de trouver les paramètres.

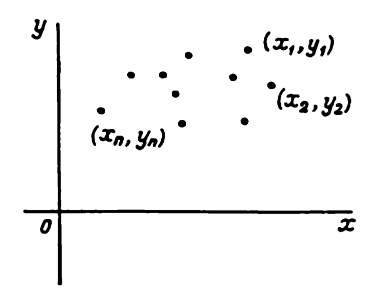


Fig. 14.7.1

Nous allons commencer par le cas le plus simple, lorsqu'on connaît d'avance, ne serait-ce qu'approximativement, la direction des axes principaux de dispersion. Le traitement se trouve alors simplifié : en orientant les axes de coordonnées parallèlement aux axes principaux de dispersion, on peut rendre les variables aléatoires pratiquement indépendantes et les traiter indépendamment. La loi de répartition des coordonnées X et Y sera caractérisée par quatre paramètres seulement: ce sont les coordonnées du centre de dispersion m_x, m_y et les écarts quadratiques moyens principaux σ_x, σ_y. Les estimations de ces paramètres sont données par les formules:

$$\left.\begin{aligned} \tilde{m}_x &= \frac{\sum_{i=1}^{n} x_i}{n}; \quad \tilde{m}_y = \frac{\sum_{i=1}^{n} y_i}{n}; \\ \tilde{\sigma}_x &= \sqrt{\frac{\sum_{i=1}^{n} (x_i - m_x)^2}{n-1}}; \\ \tilde{\sigma}_y &= \sqrt{\frac{\sum_{i=1}^{n} (y_i - m_y)^2}{n-1}}. \end{aligned}\right\} \qquad (14.7.1)$$

Passons maintenant à un cas plus compliqué, où la direction des axes principaux de dispersion est inconnue d'avance et doit également être déterminée par l'expérience. Dans ce cas il y a lieu d'estimer tous les cinq paramètres, à savoir les coordonnées du centre de

dispersion m_x, m_y, l'angle α et les écarts quadratiques moyens principaux σ_ξ, σ_η (fig. 14.7.2).

Les estimations des coordonnées du centre de dispersion sont comme dans le cas précédent données par les formules:

$$\tilde{m}_x = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \tilde{m}_y = \frac{\sum_{i=1}^{n} y_i}{n}. \qquad (14.7.2)$$

Passons à l'estimation de l'angle α. Supposons connues les directions des axes principaux de dispersion et faisons passer par le point (m_x, m_y) les axes principaux $O\xi$, $O\eta$ (fig. 14.7.2). Dans le système $\xi O\eta$ les coordonnées du point aléatoire (X, Y) seront:

$$\left.\begin{aligned} \Xi &= (X - m_x)\cos\alpha + (Y - m_y)\sin\alpha; \\ \text{H} &= -(X - m_x)\sin\alpha + (Y - m_y)\cos\alpha \end{aligned}\right\}$$

ou

$$\left.\begin{aligned} \Xi &= \overset{\circ}{X}\cos\alpha + \overset{\circ}{Y}\sin\alpha; \\ \text{H} &= -\overset{\circ}{X}\sin\alpha + \overset{\circ}{Y}\cos\alpha. \end{aligned}\right\} \qquad (14.7.3)$$

Il est évident que les variables Ξ, H auront des espérances mathématiques nulles:

$$M[\Xi] = M[\text{H}] = 0.$$

Comme $O\xi$, $O\eta$ sont les axes principaux de dispersion, les variables Ξ, H sont indépendantes. Mais des variables indépendantes normales sont également non corrélées; par conséquent il nous suffit de trouver une valeur de l'angle α, telle que les variables Ξ, H ne soient pas corrélées. Cette valeur de α donnera la direction des axes principaux de dispersion.

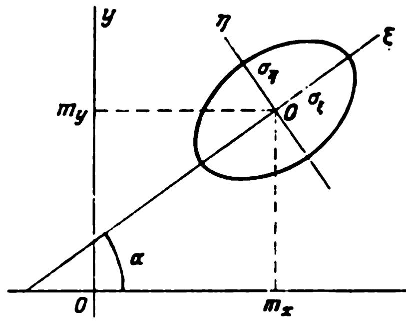


Fig. 14.7.2

Calculons la covariance des variables (Ξ, H). En multipliant les égalités (14.7.3) et en prenant l'espérance mathématique du produit on a

$$K_{\xi\eta} = M[\Xi\text{H}] = -M[\overset{\circ}{X}{}^2]\sin\alpha\cos\alpha + M[\overset{\circ}{X}\overset{\circ}{Y}](\cos^2\alpha - \sin^2\alpha) +$$
$$+ M[\overset{\circ}{Y}{}^2]\sin\alpha\cos\alpha = -\frac{1}{2}\sin 2\alpha\,(D_x - D_y) + K_{xy}\cos 2\alpha.$$

En annulant cette expression et en divisant les deux membres par $\cos 2\alpha$ on a:

$$\operatorname{tg} 2\alpha = \frac{2K_{xy}}{D_x - D_y}. \qquad (14.7.4)$$

L'équation (14.7.4) donnera deux valeurs de l'angle α: α_1 et α_2 différant de $\pi/2$. Ces deux angles déterminent la direction des axes principaux de dispersion *).

En remplaçant dans (14.7.4) les grandeurs K_{xy}, D_x, D_y par leurs estimations on obtient l'estimation de l'angle α:

$$\tilde{\alpha}=\frac{1}{2}\operatorname{arc\,tg}\frac{2\tilde{K}_{xy}}{\tilde{D}_x-\tilde{D}_y}.$$

Cherchons les estimations pour les écarts quadratiques moyens principaux σ_ξ, σ_η. A cet effet nous allons calculer les variances des variables Ξ, H données par les formules (14.7.3) en nous servant du théorème sur la variance d'une fonction linéaire:

$$D_\xi=D_x\cos^2\alpha+D_y\sin^2\alpha+2K_{xy}\sin\alpha\cos\alpha;$$
$$D_\eta=D_x\sin^2\alpha+D_y\cos^2\alpha-2K_{xy}\sin\alpha\cos\alpha,$$

d'où les estimations des variances principales:

$$\left.\begin{aligned}\tilde{D}_\xi&=\tilde{D}_x\cos^2\tilde{\alpha}+\tilde{K}_{xy}\sin 2\tilde{\alpha}+\tilde{D}_y\sin^2\tilde{\alpha};\\ \tilde{D}_\eta&=\tilde{D}_x\sin^2\tilde{\alpha}-\tilde{K}_{xy}\sin 2\tilde{\alpha}+\tilde{D}_y\cos^2\tilde{\alpha}.\end{aligned}\right\}\qquad(14.7.5)$$

Les estimations des écarts quadratiques moyens principaux sont:

$$\left.\begin{aligned}\tilde{\sigma}_\xi&=\sqrt{\tilde{D}_\xi};\\ \tilde{\sigma}_\eta&=\sqrt{\tilde{D}_\eta}.\end{aligned}\right\}\qquad(14.7.6)$$

Rassemblons toutes les formules à utiliser pour estimer, d'après les observations, les paramètres de dispersion de deux variables aléatoires (X, Y) normales dans un plan, dans le cas où l'on ne connaît pas à l'avance la direction des axes de dispersion:

$$\left.\begin{aligned}&\tilde{m}_x=\frac{\sum_{i=1}^{n}x_i}{n};\quad \tilde{m}_y=\frac{\sum_{i=1}^{n}y_i}{n};\\ &\tilde{\alpha}=\frac{1}{2}\operatorname{arc\,tg}\frac{2\tilde{K}_{xy}}{\tilde{D}_x-\tilde{D}_y};\\ &\tilde{\sigma}_\xi=\sqrt{\tilde{D}_x\cos^2\tilde{\alpha}+\tilde{K}_{xy}\sin 2\tilde{\alpha}+\tilde{D}_y\sin^2\tilde{\alpha}};\\ &\tilde{\sigma}_\eta=\sqrt{\tilde{D}_x\sin^2\tilde{\alpha}-\tilde{K}_{xy}\sin 2\tilde{\alpha}+\tilde{D}_y\cos^2\tilde{\alpha}},\end{aligned}\right\}\qquad(14.7.7)$$

*) Notons que l'expression (14.7.4) coïncide avec l'expression (9.2.2) du chapitre 9 pour l'angle α, donnant la direction des axes de symétrie de l'ellipse de dispersion.

où les estimations des variances et des covariances doivent être calculées par les formules suivantes:

$$\left.\begin{aligned} \tilde{D}_x &= \frac{\sum_{i=1}^{n}(x_i-\tilde{m}_x)^2}{n-1}; \quad \tilde{D}_y = \frac{\sum_{i=1}^{n}(y_i-\tilde{m}_y)^2}{n-1};\\ \tilde{K}_{xy} &= \frac{\sum_{i=1}^{n}(x_i-\tilde{m}_x)(y_i-\tilde{m}_y)}{n-1}, \end{aligned}\right\} \qquad (14.7.8)$$

ou, ce qui est souvent plus commode, en fonction des moments centrés par les formules:

$$\left.\begin{aligned} \tilde{D}_x &= \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n} x_i^2}{n} - \tilde{m}_x^2\right];\\ \tilde{D}_y &= \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n} y_i^2}{n} - \tilde{m}_y^2\right];\\ \tilde{K}_{xy} &= \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n} x_i y_i}{n} - \tilde{m}_x\tilde{m}_y\right]. \end{aligned}\right\} \qquad (14.7.9)$$

14.8. Lissage des courbes expérimentales par la méthode des moindres carrés

Le problème du lissage des courbes expérimentales est très voisin des questions liées au traitement des données expérimentales étudiées dans le présent chapitre.

Supposons qu'une expérience ait pour but d'étudier la relation entre une certaine grandeur physique y et une autre grandeur physique x (par exemple, le chemin parcouru par un corps en fonction du temps; la vitesse initiale d'un obus en fonction de la température de la charge; la portance en fonction de l'angle d'attaque, etc.). On suppose que les grandeurs x et y sont liées par la relation fonctionnelle

$$y = \varphi(x). \qquad (14.8.1)$$

C'est cette relation qu'il y a lieu de trouver à partir des données expérimentales.

Supposons qu'on ait obtenu expérimentalement une série de points et tracé le graphique de y en fonction de x (fig. 14.8.1). Généralement les points expérimentaux ne forment pas une courbe régulière

mais présentent une certaine « dispersion », donc s'écartent d'une façon aléatoire de la relation générale perceptible. Ces écarts sont liés aux *erreurs de mesure* inévitables dans toute expérience.

La question qui se pose est: comment à partir de ces données expérimentales trouver la relation existant entre y et x et ceci avec la meilleure approximation possible?

On sait que par n points quelconques de coordonnées (x_i, y_i) on peut toujours mener une courbe s'exprimant analytiquement par

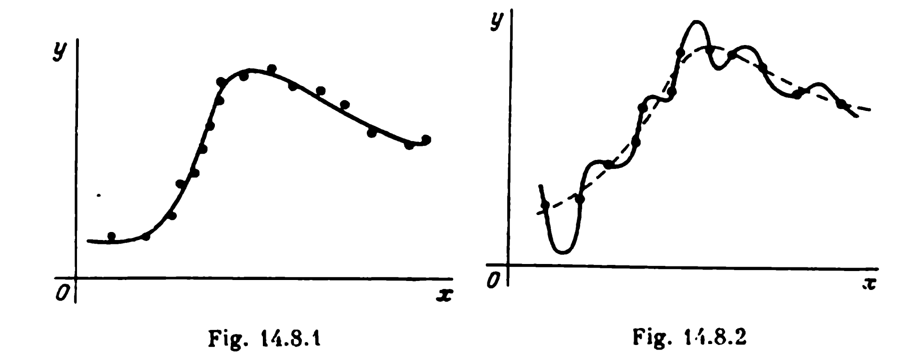


Fig. 14.8.1 Fig. 14.8.2

un polynôme de degré $(n - 1)$ et passant exactement par chacun des points (fig. 14.8.2). Cependant une telle solution du problème n'est pas en général satisfaisante, car presque toujours la non-régularité des points expérimentaux, comme par exemple sur les figures 14.8.1 et 14.8.2, est conditionnée non pas par la relation existant entre x et y mais uniquement par les erreurs de mesure. On peut facilement s'en rendre compte en comparant les écarts observés (dispersion des points) avec les erreurs éventuelles des appareils de mesure utilisés.

Très souvent, dans la pratique on se trouve devant le problème du *lissage* des courbes expérimentales. Les données expérimentales doivent être traitées de façon à exprimer le plus exactement possible la tendance générale qui se dessine dans la relation entre y et x, tout en lissant les écarts aléatoires injustifiés, dus aux erreurs inévitables de l'expérience même.

Pour résoudre ce problème on utilise généralement la « méthode des moindres carrés ». Cette méthode permet, pour une forme donnée de la fonction $y = \varphi(x)$, de choisir ses paramètres numériques de telle sorte que la courbe $y = \varphi(x)$ corresponde de la meilleure dans un certain sens façon aux données expérimentales.

Nous allons dire quelques mots sur la manière de choisir la forme de la courbe $y = \varphi(x)$. Souvent ceci se voit immédiatement en examinant la relation expérimentale obtenue. Par exemple, les points

expérimentaux représentés sur la figure 14.8.3 laissent indubitablement penser à une relation linéaire de la forme $y = ax + b$. La relation apparaissant sur la figure 14.8.4 peut être exprimée par un polynôme du second degré de la forme $y = ax^2 + bx + c$. Dans le cas

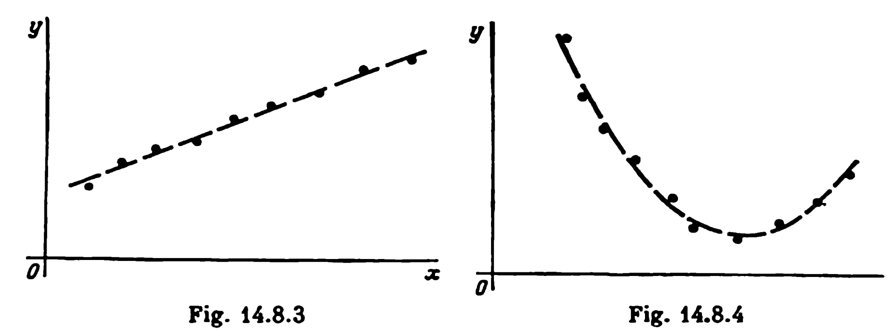


Fig. 14.8.3 Fig. 14.8.4

d'une fonction périodique, on peut souvent utiliser plusieurs harmoniques de la série trigonométrique, etc.

Souvent la forme de la relation (linéaire, quadratique, exponentielle, etc.) est donnée par des considérations physiques, et il y a lieu de trouver expérimentalement certains *paramètres* de cette relation.

Tableau 14.8.1

i	x_i	y_i
1	x_1	y_1
2	x_2	y_2
...	...	...
i	x_i	y_i
...	...	...
n	x_n	y_n

Dans ce paragraphe nous allons étudier la manière dont on peut choisir les paramètres d'une fonction de forme donnée à partir des données expérimentales.

Soient les résultats des n expériences indépendantes disposés dans un tableau statistique simple (tableau 14.8.1), où i est le numéro d'ordre de l'expérience; x_i la valeur de l'argument et y_i la valeur correspondante de la fonction.

Les points (x_i, y_i) sont portés sur un graphique (fig. 14.8.5).

Des considérations théoriques ou autres ont permis de déterminer la forme de la relation $y = \varphi(x)$. La fonction $y = \varphi(x)$ contient plusieurs paramètres numériques $a, b, c, \ldots$ Ces paramètres doivent être choisis de telle sorte que la courbe exprime le mieux, en un certain sens, la relation obtenue expérimentalement.

Comme dans tout problème de lissage il y a lieu d'établir tout d'abord ce que nous considérerons être la meilleure solution. Par exemple, on peut considérer la meilleure la solution assurant le minimum des écarts entre la courbe approchée et des points expérimentaux ou bien on peut exiger que soit minimale la somme des valeurs absolues des écarts entre les points et la courbe, etc. Pour chacune de ces exigences on obtient une solution particulière du problème et certaines valeurs des paramètres a, b, c . . . Cependant, la plus employée pour la solution de ces problèmes est la *méthode des moindres carrés*, qui consiste à rendre *minimale la somme des carrés des écarts entre les points expérimentaux et la courbe de lissage*. La méthode des moindres carrés présente par rapport aux autres méthodes de lissage d'avantages importants: premièrement, elle permet une détermination des paramètres a, b, c, . . . assez simple, de plus elle repose sur un fondement théorique probabiliste sérieux.

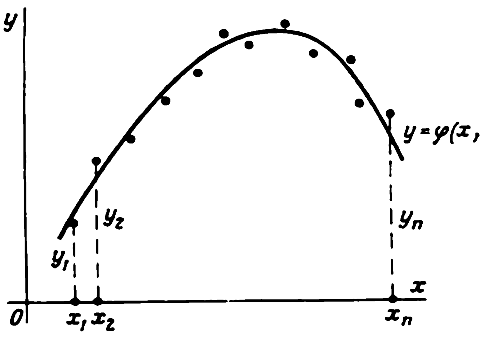


Fig. 14.8.5

Examinons un peu plus près ce fondement. Soit $y = \varphi(x)$ l'expression exacte de la relation réelle existant entre y et x. Les points expérimentaux s'écartent de cette relation par suite des *erreurs* inévitables *de mesure*. Nous avons déjà mentionné que les erreurs de mesure suivent généralement la loi normale. Supposons qu'il en soit ainsi. Considérons une valeur quelconque x_i de l'argument. Le résultat de l'expérience est la variable aléatoire Y_i répartie suivant une loi normale d'espérance mathématique $\varphi(x_i)$ et d'écart quadratique moyen σ_i, ce dernier caractérisant l'erreur de mesure. Supposons que l'erreur de mesure soit la même en tous les points:

$$\sigma_1 = \sigma_2 = \ldots = \sigma_n = \sigma.$$

La loi normale de répartition de la variable Y peut alors s'écrire comme suit:

$$f_i(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y_i - \varphi(x_i)]^2}{2\sigma^2}}. \qquad (14.8.2)$$

Supposons que l'expérience qui est dans notre exemple une série de mesures ait réalisé l'événement consistant en ce que les variables aléatoires $(Y_1, Y_2, \ldots, Y_n)$ ont pris les valeurs $(y_1, y_2, \ldots, y_n)$. Il y a lieu de choisir les espérances mathématiques $\varphi(x_1)$, $\varphi(x_2)$, . . .

. . ., $\varphi(x_n)$ de telle sorte que la *probabilité de cet événement soit maximale* *).

En toute rigueur la probabilité de l'un quelconque des événements $Y_i = y_i$, tout comme celle de leur réalisation simultanée, est égale à zéro car les variables Y_i sont continues; c'est pourquoi nous allons nous servir non pas de la probabilité des événements $Y_i = y_i$ mais des *éléments de probabilité* correspondants:

$$f_i(y_i)\,dy_i = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{[y_i-\varphi(x_i)]^2}{2\sigma^2}}\, dy_i. \qquad (14.8.3)$$

Nous allons calculer la probabilité pour que le système de variables aléatoires $Y_1, Y_2, \ldots, Y_n$ prenne l'ensemble des valeurs se trouvant chacune dans l'intervalle

$$(y_i,\ y_i + dy_i) \quad (i = 1, 2, \ldots, n).$$

Les expériences étant indépendantes, cette probabilité est égale au produit des éléments de probabilité (14.8.3) pour tous les i:

$$\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{[y_i-\varphi(x_i)]^2}{2\sigma^2}}\, dy_i = K e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}[y_i-\varphi(x_i)]^2}, \qquad (14.8.4)$$

où K est un coefficient dépendant de $\varphi(x_i)$.

Il y a lieu de choisir les espérances mathématiques $\varphi(x_1)$, $\varphi(x_2), \ldots, \varphi(x_n)$ de telle sorte que l'expression (14.8.4) soit maximale.

La grandeur

$$e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}[y_i-\varphi(x_i)]^2}$$

est toujours inférieure à l'unité, donc sa valeur est maximale lorsque l'exposant est minimal en valeur absolue:

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i-\varphi(x_i)]^2 = \min.$$

Ceci donne, le facteur constant $\frac{1}{2\sigma^2}$ mis à part, la condition imposée par la méthode des moindres carrés: pour que l'ensemble de valeurs observées

$$y_1, y_2, \ldots, y_n.$$

soit *le plus probable* il faut choisir la fonction $\varphi(x)$ telle que *la somme des carrés des écarts des valeurs observées y_i par rapport à $\varphi(x_i)$ soit*

*) C'est le principe dit « du maximum de vraisemblance ».

minimale :

$$\sum_{i=1}^{n} [y_i - \varphi(x_i)]^2 = \min.$$

Nous avons ainsi justifié la méthode des moindres carrés, en nous basant sur la loi normale des erreurs de mesure et la condition de probabilité maximale d'un ensemble donné d'erreurs.

Passons au calcul des paramètres a, b, c . . . partant du principe des moindres carrés. Supposons que l'on ait un tableau de données expérimentales (tableau 14.8.1) et que certaines considérations : physiques, théoriques ou simplement l'allure de la relation observée, aient permis de choisir la forme générale de la fonction $y = \varphi(x)$ dépendant de plusieurs paramètres numériques $a, b, c, \ldots$. Il y a lieu de déterminer ces paramètres à l'aide de la méthode des moindres carrés de telle sorte que la somme des carrés des écarts y_i par rapport à $\varphi(x_i)$ soit minimale. Ecrivons y comme suit :

$$y = \varphi(x;\ a,\ b,\ c,\ \ldots) \tag{14.8.5}$$

Le choix des paramètres a, b, c, . . . est régi par la condition

$$\sum_{i=1}^{n} [y_i - \varphi(x_i;\ a,\ b,\ c,\ \ldots)]^2 = \min. \tag{14.8.6}$$

Nous allons trouver les valeurs de a, b, c, . . . rendant minimal le premier membre de l'expression (14.8.6). A cet effet dérivons-le par rapport à a, b, c, et annulons les dérivées :

$$\left.\begin{aligned} &\sum_{i=1}^{n} [y_i - \varphi(x_i;\ a,\ b,\ c,\ \ldots)] \left(\frac{\partial\varphi}{\partial a}\right)_i = 0, \\ &\sum_{i=1}^{n} [y_i - \varphi(x_i;\ a,\ b,\ c,\ \ldots)] \left(\frac{\partial\varphi}{\partial b}\right)_i = 0, \\ &\sum_{i=1}^{n} [y_i - \varphi(x_i;\ a,\ b,\ c,\ \ldots)] \left(\frac{\partial\varphi}{\partial c}\right)_i = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \tag{14.8.7}$$

où $\left(\frac{\partial\varphi}{\partial a}\right)_i = \varphi'_a(x_i;\ a,\ b,\ c,\ \ldots)$ est la dérivée partielle de la fonction φ par rapport au paramètre a au point x_i et $\left(\frac{\partial\varphi}{\partial b}\right)_i$, $\left(\frac{\partial\varphi}{\partial c}\right)_i$, . . . les dérivées partielles par rapport à b, c, ...

Le système (14.8.7) contient autant d'équations que d'inconnues a, b, c, ...

Le système (14.8.7) ne peut être résolu sous une forme générale ; pour cela il faudrait se donner la forme de la fonction φ.

Nous allons envisager deux cas que l'on rencontre souvent dans la pratique; le cas où φ est une fonction linéaire et lorsque c'est un polynôme du second degré (parabole).

1. Choix des paramètres d'une fonction linéaire par la méthode des moindres carrés

L'expérience a fourni l'ensemble des valeurs (x_i, y_i) $(i = 1, 2, \ldots, n$; fig. 14.8.6).

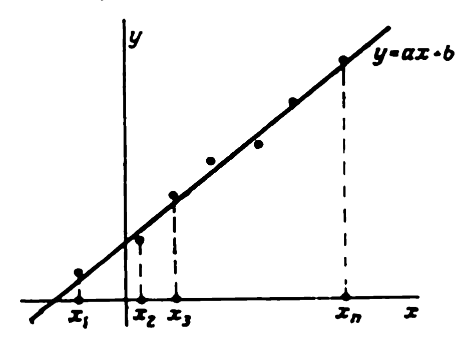


Fig. 14.8.6

Déterminer par la méthode des moindres carrés les paramètres a, b de la fonction linéaire

$$y = ax + b,$$

représentant la relation expérimentale étudiée.

S o l u t i o n: On a:

$$y = \varphi(x;\ a,\ b) = ax + b. \tag{14.8.8}$$

En dérivant l'expression (14.8.8) par rapport à a et b on a:

$$\frac{\partial\varphi}{\partial a} = x; \quad \left(\frac{\partial\varphi}{\partial a}\right)_i = x_i;$$

$$\frac{\partial\varphi}{\partial b} = 1; \quad \left(\frac{\partial\varphi}{\partial b}\right)_i = 1.$$

En substituant les résultats dans les formules (14.8.7) on obtient deux équations permettant de calculer a et b:

$$\sum_{i=1}^{n} [y_i - (ax_i + b)]\, x_i = 0;$$

$$\sum_{i=1}^{n} [y_i - (ax_i + b)] = 0,$$

ou en enlevant les parenthèses et réalisant la sommation

$$\left.\begin{aligned} \sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i^2 - b \sum_{i=1}^{n} x_i = 0, \\ \sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i - bn = 0. \end{aligned}\right\} \qquad (14.8.9)$$

Divisons les deux équations (14.8.9) par n:

$$\left.\begin{aligned} \frac{\sum_{i=1}^{n} x_i y_i}{n} - a \frac{\sum_{i=1}^{n} x_i^2}{n} - b \frac{\sum_{i=1}^{n} x_i}{n} = 0, \\ \frac{\sum_{i=1}^{n} y_i}{n} - a \frac{\sum_{i=1}^{n} x_i}{n} - b = 0. \end{aligned}\right\} \qquad (14.8.10)$$

Les sommes entrant dans l'équation (14.8.10) ne sont rien d'autre que les *moments statistiques*:

$$\frac{\sum_{i=1}^{n} x_i}{n} = m_x^*; \quad \frac{\sum_{i=1}^{n} x_i^2}{n} = \alpha_2^* [X];$$

$$\frac{\sum_{i=1}^{n} y_i}{n} = m_y^*; \quad \frac{\sum_{i=1}^{n} x_i y_i}{n} = \alpha_{1,1}^* [X, Y].$$

En substituant ces équations dans le système (14.8.10) on obtient:

$$\left.\begin{aligned} \alpha_{1,1}^* [X, Y] - a\alpha_2^* [X] - bm_x^* = 0, \\ m_y^* - am_x^* - b = 0. \end{aligned}\right\} \qquad (14.8.11)$$

En tirant b de la seconde équation (14.8.11) et en le substituant dans la première on a:

$$b = m_y^* - am_x^*$$

$$\alpha_{1,1}^* [X, Y] - a\alpha_2^* [X] - (m_y^* - am_x^*) m_x^* = 0.$$

Résolvons cette dernière équation par rapport à a:

$$a = \frac{\alpha_{1,1}^* [X, Y] - m_x^* m_y^*}{\alpha_2^* [X] - (m_x^*)^2} \qquad (14.8.12)$$

L'expression (14.8.12) peut être simplifiée par l'introduction des moments centrés au lieu des moments initiaux. En effet:

$$\alpha_{1,1}^* [X, Y] - m_x^* m_y^* = K_{xy}^*,$$

$$\alpha_2^* [X] - (m_2^*)^2 = D_x^*,$$

d'où

$$a = \frac{K^*_{xy}}{D^*_x}; \quad b = m^*_y - am^*_x, \tag{14.8.13}$$

avec

$$\left.\begin{aligned} m^*_x &= \frac{\sum_{i=1}^{n} x_i}{n}; \quad m^*_y = \frac{\sum_{i=1}^{n} y_i}{n}; \\ K^*_{xy} &= \frac{\sum_{i=1}^{n} (x_i - m^*_x)(y_i - m^*_y)}{n}; \\ D^*_x &= \frac{\sum_{i=1}^{n} (x_i - m^*_x)^2}{n}. \end{aligned}\right\} \tag{14.8.14}$$

Nous avons résolu le problème posé et trouvé la relation linéaire entre y et x sous la forme:

$$y = \frac{K^*_{xy}}{D^*_x} x + m^*_y - \frac{K^*_{xy}}{D^*_x} m^*_x,$$

où en transportant m^*_y dans le premier membre:

$$y - m^*_y = \frac{K^*_{xy}}{D^*_x} (x - m^*_x). \tag{14.8.15}$$

Nous avons exprimé les coefficients de la relation linéaire en fonction des moments centrés et non des moments initiaux, seulement parce que sous cette forme les formules sont plus simples. Dans les applications pratiques il peut s'avérer plus commode de calculer K^*_{xy} et D^*_x non pas par les formules (14.8.14) mais en fonction des moments initiaux d'ordre 2:

$$\left.\begin{aligned} K^*_{xy} &= \frac{\sum_{i=1}^{n} x_i y_i}{n} - m^*_x m^*_y, \\ D^*_x &= \frac{\sum_{i=1}^{n} x_i^2}{n} - (m^*_x)^2. \end{aligned}\right\} \tag{14.8.16}$$

Pour que les formules (14.8.16) ne conduisent pas à des différences de nombres voisins, il y a lieu de reporter l'origine en un point pas trop éloigné des espérances mathématiques m^*_x, m^*_y.

2. Calcul des paramètres d'une parabole du second degré par la méthode des moindres carrés

L'expérience a fourni l'ensemble des valeurs (x_i, y_i) $(i = 1, 2, \ldots, n$; fig. 14.8.7). Trouver par la méthode des moindres carrés

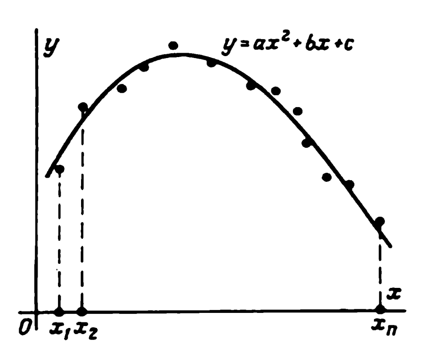


Fig. 14.8.7

les paramètres d'une fonction quadratique (parabole du second degré) de la forme:

$$y = ax^2 + bx + c$$

correspondant à la relation observée expérimentalement. On a:

$$y = \varphi(x;\ a,\ b,\ c) = ax^2 + bx + c,$$

$$\frac{\partial\varphi}{\partial a} = x^2; \qquad \left(\frac{\partial\varphi}{\partial a}\right)_i = x_i^2;$$

$$\frac{\partial\varphi}{\partial b} = x; \qquad \left(\frac{\partial\varphi}{\partial b}\right)_i = x_i;$$

$$\frac{\partial\varphi}{\partial c} = 1; \qquad \left(\frac{\partial\varphi}{\partial c}\right)_i = 1.$$

En substituant les résultats ci-dessus dans l'équation (14.8.7) on obtient:

$$\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]\, x_i^2 = 0,$$

$$\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]\, x_i = 0,$$

$$\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)] = 0,$$

ou après transformations :

$$\left.\begin{aligned}
&\frac{\sum\limits_{i=1}^{n} x_i^2 y_i}{n} - a\frac{\sum\limits_{i=1}^{n} x_i^4}{n} - b\frac{\sum\limits_{i=1}^{n} x_i^3}{n} - c\frac{\sum\limits_{i=1}^{n} x_i^2}{n} = 0,\\
&\frac{\sum\limits_{i=1}^{n} x_i y_i}{n} - a\frac{\sum\limits_{i=1}^{n} x_i^3}{n} - b\frac{\sum\limits_{i=1}^{n} x_i^2}{n} - c\frac{\sum\limits_{i=1}^{n} x_i}{n} = 0,\\
&\frac{\sum\limits_{i=1}^{n} y_i}{n} - a\frac{\sum\limits_{i=1}^{n} x_i^2}{n} - b\frac{\sum\limits_{i=1}^{n} x_i}{n} - c = 0.
\end{aligned}\right\} \qquad (14.8.17)$$

Les coefficients de ce système sont également les moments statistiques du couple X, Y :

$$\frac{\sum\limits_{i=1}^{n} x_i}{n} = m_x^* = \alpha_1^*[X]; \qquad \frac{\sum\limits_{i=1}^{n} y_i}{n} = m_y^* = \alpha_1^*[Y];$$

$$\frac{\sum\limits_{i=1}^{n} x_i^2}{n} = \alpha_2^*[X]; \qquad \frac{\sum\limits_{i=1}^{n} x_i^3}{n} = \alpha_3^*[X]; \qquad \frac{\sum\limits_{i=1}^{n} x_i^4}{n} = \alpha_4^*[X];$$

$$\frac{\sum\limits_{i=1}^{n} x_i y_i}{n} = \alpha_{1,1}^*[X, Y]; \qquad \frac{\sum\limits_{i=1}^{n} x_i^2 y_i}{n} = \alpha_{2,1}^*[X, Y].$$

En utilisant ces expressions on peut écrire le système d'équations (14.8.7) sous une forme assez simple. En effet, compte tenu de ce que $\alpha_0^*[X] = 1$; $\alpha^*_{0,1}[X, Y] = \alpha_1^*[Y]$ et transférant les termes ne contenant pas d'inconnues dans les seconds membres on obtient :

$$\left.\begin{aligned}
&\alpha_4^*[X]\,a + \alpha_3^*[X]\,b + \alpha_2^*[X]\,c = \alpha_{2,1}^*[X, Y],\\
&\alpha_3^*[X]\,a + \alpha_2^*[X]\,b + \alpha_1^*[X]\,c = \alpha_{1,1}^*[X, Y],\\
&\alpha_2^*[X]\,a + \alpha_1^*[X]\,b + \alpha_0^*[X]\,c = \alpha_{0,1}^*[X, Y].
\end{aligned}\right\} \qquad (14.8.18)$$

Il est facile de trouver la loi donnant les coefficients dans les équations (14.8.18) : le premier membre contient dans l'ordre décroissant seuls les moments de la variable X, le second membre représente les moments du couple (X, Y), de plus, les moments en X vont en ordre décroissant d'une équation à l'autre, et ceux en Y sont toujours du premier ordre *).

*) La solution du système (14.8.18) par rapport aux inconnues a, b, c n'est pas donnée ici sous une forme générale, celle-ci étant trop compliquée ; en pratique on ne se sert pas en général de la méthode des déterminants, mais on procède par exclusion successive des inconnues.

Les coefficients d'une parabole d'ordre quelconque seront donnés par des équations semblables.

On voit que dans le cas où une relation expérimentale est lissée d'après la méthode des moindres carrés par un polynôme d'ordre quelconque, les coefficients de ce polynôme sont les solutions d'un système d'équations linéaires. Les coefficients de ce système d'équations linéaires sont les moments statistiques d'ordre quelconque, caractérisant le système de grandeurs (X, Y), celui-ci étant considéré comme un système de variables aléatoires.

La solution du problème du lissage d'une relation expérimentale par la méthode des moindres carrés est presque aussi simple lorsque la fonction de lissage est non pas un polynôme mais une somme de fonctions arbitraires données $\varphi_1(x), \varphi_2(x), \ldots, \varphi_k(x)$ avec les coefficients $a_1, a_2, \ldots, a_k$:

$$y = \varphi(x; a_1, a_2, \ldots, a_k) = a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_k\varphi_k(x) = \sum_{i=1}^{k} a_i\varphi_i(x) \quad (14.8.19)$$

et qu'il y a lieu de trouver les coefficients a_i.

Par exemple, on peut employer à cette fin un polynôme trigonométrique:

$$\varphi(x; a_1, a_2, a_3, a_4) = a_1 \cos \omega x + a_2 \sin \omega x + a_3 \cos 2\omega x + a_4 \sin 2\omega x$$

ou une combinaison linéaire de fonctions exponentielles:

$$\varphi(x; a_1, a_2, a_3) = a_1 e^{\alpha t} + a_2 e^{\beta t} + a_3 e^{\gamma t}, \text{ etc.}$$

Dans le cas où la fonction est donnée par une expression du type (14.8.19), les coefficients $a_1, a_2, \ldots, a_k$ sont les solutions d'un système de k équations linéaires du type:

$$\sum_{i=1}^{n} \{y_i - [a_1\varphi_1(x_i) + a_2\varphi_2(x_i) + \ldots + a_k\varphi_k(x_i)]\}\, \varphi_1(x_i) = 0,$$

$$\sum_{i=1}^{n} \{y_i - [a_1\varphi_1(x_i) + a_2\varphi_2(x_i) + \ldots + a_k\varphi_k(x_i)]\}\, \varphi_2(x_i) = 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{i=1}^{n} \{y_i - [a_1\varphi_1(x_i) + a_2\varphi_2(x_i) + \ldots + a_k\varphi_k(x_i)]\}\, \varphi_k(x_i) = 0.$$

En prenant la somme terme à terme on a:

$$a_1 \sum_{i=1}^{n} [\varphi_1(x_i)]^2 + a_2 \sum_{i=1}^{n} \varphi_2(x_i)\varphi_1(x_i) + \ldots +$$
$$+ a_k \sum_{i=1}^{n} \varphi_k(x_i)\varphi_1(x_i) = \sum_{i=1}^{n} y_i\varphi_1(x_i).$$
$$a_1 \sum_{i=1}^{n} \varphi_1(x_i)\varphi_2(x_i) + a_2 \sum_{i=1}^{n} [\varphi_2(x_i)]^2 + \ldots +$$
$$+ a_k \sum_{i=1}^{n} \varphi_k(x_i)\varphi_2(x_i) = \sum_{i=1}^{n} y_i\varphi_2(x_i),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_1 \sum_{i=1}^{n} \varphi_i(x_i)\varphi_k(x_i) + a_2 \sum_{i=1}^{n} \varphi_2(x_i)\varphi_k(x_i) + \ldots +$$
$$+ a_k \sum_{i=1}^{n} [\varphi_k(x_i)]^2 = \sum_{i=1}^{n} y_i\varphi_k(x_i),$$

ou sous une forme plus compacte:

$$\left.\begin{array}{l} \sum_{j=1}^{k} a_j \sum_{i=1}^{n} \varphi_1(x_i)\varphi_j(x_i) = \sum_{i=1}^{n} y_i\varphi_1(x_i), \\ \sum_{j=1}^{k} a_j \sum_{i=1}^{n} \varphi_2(x_i)\varphi_j(x_i) = \sum_{i=1}^{n} y_i\varphi_2(x_i), \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{j=1}^{k} a_j \sum_{i=1}^{n} \varphi_k(x_i)\varphi_j(x_i) = \sum_{i=1}^{n} y_i\varphi_k(x_i). \end{array}\right\} \qquad (14.8.20)$$

On peut toujours résoudre le système d'équations (14.8.20) et trouver ainsi les coefficients $a_1, a_2, \ldots, a_k$.

Le problème du lissage par la méthode des moindres carrés devient beaucoup plus compliqué lorsque l'expression de la fonction $y = \varphi(x; a, b, c, \ldots)$ n'est pas linéaire par rapport aux coefficients numériques $a, b, c, \ldots$ Dans ce cas les calculs deviennent très vite inextricables. Cependant même dans ce cas souvent on arrive à obtenir la solution du problème d'une manière assez simple.

Nous allons illustrer l'idée fondamentale des procédés employés sur l'exemple le plus simple où la fonction comporte un seul paramètre non linéaire a (par exemple $y = e^{-ax^2}$ ou $y = \sin ax$ ou encore $y = 1/ax$). Sous la forme générale:

$$y = \varphi(x, a), \qquad (14.8.21)$$

a est le paramètre qu'il y a lieu de choisir par la méthode des moindres carrés pour un meilleur lissage de la relation expérimentale montrée sur la figure 14.8.8.

Nous allons résoudre ce problème de la manière suivante. Soit un certain nombre de valeurs du paramètre a, pour chacune desquelles nous allons trouver la somme des carrés des écarts $y_i - \varphi(x_i, a)$, qui est une certaine fonction de a et que nous désignerons par $\Sigma(a)$:

$$\Sigma(a) = \sum_{i=1}^{n} [y_i - \varphi(x_i, a)]^2.$$

Construisons la courbe $\Sigma(a)$ (fig. 14.8.9).

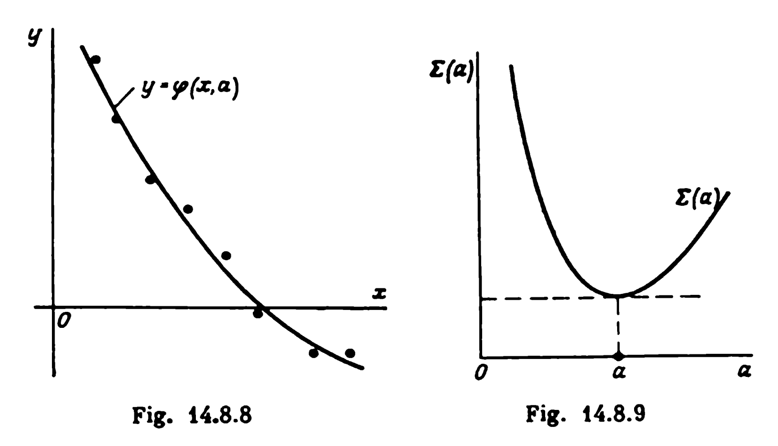


Fig. 14.8.8 Fig. 14.8.9

La valeur de a pour laquelle la courbe $\Sigma(a)$ a un minimum est adoptée pour valeur convenable du paramètre a dans l'expression (14.8.21).

On peut en principe d'une manière analogue, sans résoudre les équations (14.8.7), choisir conjointement deux paramètres (a, b)

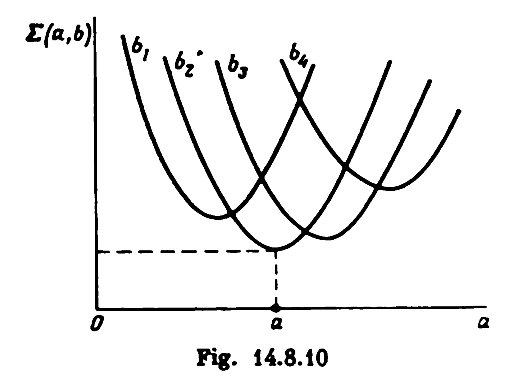


Fig. 14.8.10

satisfaisant au principe des moindres carrés; le problème ne s'en trouve que peu compliqué et revient au tracé non pas d'un seul, mais de plusieurs graphiques (fig. 14.8.10); il y aura lieu alors de chercher

l'ensemble des valeurs a, b, donnant le minimum absolu à la somme des carrés des écarts Σ (a, b).

Exemple 1. Dans une expérience on étudie la relation entre la profondeur de pénétration h d'un corps dans un obstacle et l'énergie spécifique $\mathcal{E}$

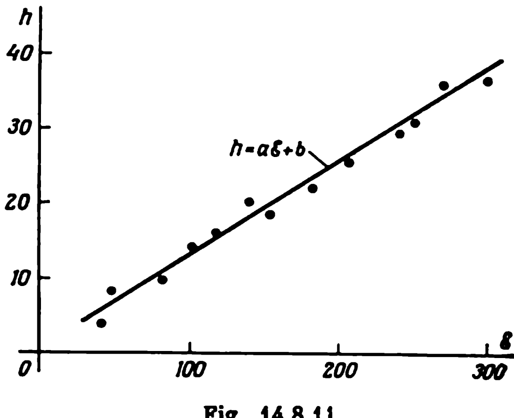


Fig. 14.8.11

(énergie par un cm carré de surface d'impact). Les données expérimentales sont consignées au tableau 14.8.2 et représentées graphiquement (fig. 14.8.11).

Tableau 14.8.2

i	$\mathcal{E}_i$, kgm/cm²	h_i, mm	i	$\mathcal{E}_i$, kgm/cm²	h_i, mm
1	41	4	8	180	23
2	50	8	9	208	26
3	81	10	10	241	30
4	104	14	11	250	31
5	120	16	12	269	36
6	139	20	13	301	37
7	154	19			

Trouver et construire par la méthode des moindres carrés la droite donnant la relation entre h et $\mathcal{E}$.

Solution. On a

$$m_{\mathcal{E}}^* = \frac{\sum_i \mathcal{E}_i}{13} = \frac{2137}{13} \approx 164{,}4, \quad m_h^* = \frac{\sum_i h_i}{13} = \frac{274}{13} \approx 21{,}1.$$

Pour le traitement à l'aide des moments initiaux transportons l'origine des coordonnées dans un point voisin de la moyenne:

$$\mathcal{E}_0 = 150; \quad h_0 = 20.$$

Dressons un nouveau tableau pour les valeurs des grandeurs:

$$\mathscr{E}' = \mathscr{E} - \mathscr{E}_0 ; \quad h' = h - h_0$$

(tableau 14.8.3).

Tableau 14.8.3

i	$\mathscr{E}'_i = \mathscr{E}_i - 150$	$h'_i = h_i - 20$	i	$\mathscr{E}'_i = \mathscr{E}_i - 150$	$h'_i = h_i - 20$
1	−109	−16	8	30	3
2	−100	−12	9	58	6
3	−69	−10	10	91	10
4	−46	−6	11	100	11
5	−30	−4	12	119	16
6	−11	0	13	151	17
7	4	−1			

Calculons les moments:

$$\alpha_2^* [\mathscr{E}'] = \frac{\sum_i \mathscr{E}_i'^2}{13} \approx 6869 ;$$

$$D_{\mathscr{E}}^* = 6869 - (m_{\mathscr{E}'}^*)^2 = 6869 - (164,4 - 150)^2 = 6662 ;$$

$$\alpha_{1,1}^* [\mathscr{E}', h'] = \frac{\sum_i \mathscr{E}_i' h_i'}{13} \approx 842 ;$$

$$K_{\mathscr{E}h}^* = \alpha_{1,1}^* [\mathscr{E}', h'] - m_{\mathscr{E}'}^* m_{h'}^* = 842 - (164,4 - 150)(21,1 - 20) \approx 842 - 16 = 826.$$

L'équation cherchée de la droite est:

$$h - m_h^* = \frac{K_{\mathscr{E}h}^*}{D_{\mathscr{E}}^*} (\mathscr{E} - m_{\mathscr{E}}^*)$$

ou

$$h - 21,1 = 1,124 (\mathscr{E} - 164,4). \qquad (14.8.22)$$

La droite (14.8.22) est représentée sur la figure 14.8.11.

Exemple 2. Un condensateur chargé à une tension de $U_0 = 100$ volts se décharge sur une résistance. On mesure la tension U entre les armatures du condensateur toutes les secondes pendant un intervalle de temps de 10 secondes. La tension est mesurée à 5 volts près. Les résultats des mesures sont donnés dans le tableau 14.8.4.

Tableau 14.8.4

i	t_i, s	U_i, V	i	t_i, s	U_i, V
1	0	100	7	6	15
2	1	75	8	7	10
3	2	55	9	8	10
4	3	40	10	9	5
5	4	30	11	10	5
6	5	20			

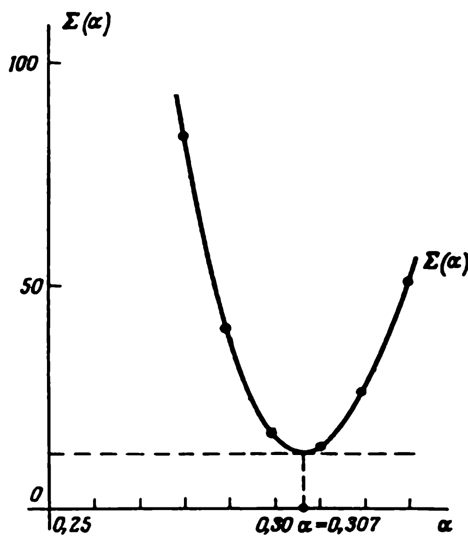


Fig. 14.8.12

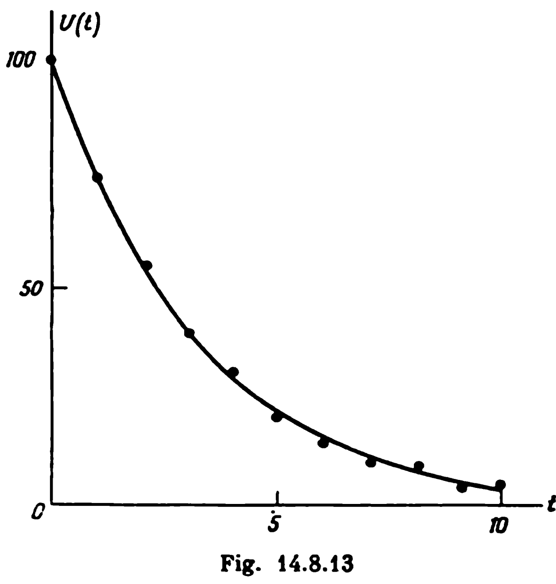


Fig. 14.8.13

Théoriquement, la tension en fonction du temps est:

$$U = U_0 e^{-\alpha t}.$$

A partir des données expérimentales choisir par la méthode des moindres carrés la valeur du paramètre α.

S o l u t i o n. Dans les tables de la fonction e^{-x} on trouve que e^{-x} est environ 0,05 pour $x \approx 3$; par conséquent, le coefficient α doit être de l'ordre de 0,3. On se donne plusieurs valeurs de α voisines de 0,3:

$$\alpha = 0{,}28;\ 0{,}29;\ 0{,}30;\ 0{,}31;\ 0{,}32;\ 0{,}33$$

et on calcule les valeurs correspondantes de la fonction:

$$U = U_0 e^{-\alpha t}$$

aux points t_i (tableau 14.8.5). Dans la dernière ligne du tableau 14.8.6 on trouve les valeurs des sommes des carrés des écarts Σ (α) en fonction de α.

Tableau 14.8.5

i	t_i	$\alpha = 0{,}28$	$\alpha = 0{,}29$	$\alpha = 0{,}30$	$\alpha = 0{,}31$	$\alpha = 0{,}32$	$\alpha = 0{,}33$
1	0	100,0	100,0	100,0	100,0	100,0	100,0
2	1	75,5	74,8	74,1	73,3	72,6	71,9
3	2	57,1	56,0	54,9	53,8	52,7	51,7
4	3	43,2	41,9	40,7	39,5	38,3	37,2
5	4	32,6	31,3	30,1	28,9	27,8	26,7
6	5	24,6	23,5	22,3	21,2	20,2	19,2
7	6	18,6	17,6	16,5	15,6	14,7	13,8
8	7	14,1	13,1	12,2	11,4	10,6	9,9
9	8	10,7	9,8	9,1	8,4	7,7	7,1
10	9	8,0	7,4	6,7	6,1	5,6	5,1
11	10	6,1	5,5	5,0	4,5	4,1	3,7
	Σ (α)	83,3	40,3	17,4	13,6	25,7	51,4

La figure 14.8.12 montre la courbe de la fonction Σ (α). On voit que le minimum est atteint à $\alpha \approx 0{,}307$. Ainsi, la méthode des moindres carrés fournit en tant que la meilleure approximation des données expérimentales la fonction $U = U_0 e^{-0{,}307t}$. Sur la figure 14.8.13 on trouvera le graphique de cette fonction ainsi que les données expérimentales.

CHAPITRE 15

NOTIONS DE BASE DE LA THÉORIE DES FONCTIONS ALÉATOIRES

15.1. Notion de fonction aléatoire

Dans notre cours de théorie des probabilités, nous avons jusqu'à présent étudié des variables aléatoires. Dans une expérience une variable aléatoire prend une certaine valeur unique, bien qu'inconnue d'avance. Par exemple: abscisse d'un coup au tir; erreur d'un radiotélémètre lors d'une seule mesure de la distance, etc.

En nous limitant à l'étude de telles variables aléatoires, nous avons vu les phénomènes aléatoires d'une manière statique, dans les conditions constantes d'une seule expérience.

Cette méthode élémentaire est souvent insuffisante pour la résolution de problèmes pratiques. Souvent on a affaire à des grandeurs aléatoires variant d'une manière continue durant l'expérience. A titre d'exemple, on peut citer l'erreur d'un radiotélémètre lors de la mesure continue d'une distance variable.

De telles grandeurs aléatoires variant durant l'expérience, à la différence des variables aléatoires ordinaires, sont appelées *fonctions aléatoires*.

La théorie des fonctions aléatoires (ou théorie des processus aléatoires ou stochastiques) a pour objet l'étude de tels phénomènes aléatoires. Cette science peut, d'une manière imagée, être appelée « dynamique des variables aléatoires ».

La théorie des fonctions aléatoires est une nouvelle branche de la théorie des probabilités qui est en voie de développement depuis quelque 30 ans. A l'heure actuelle cette théorie continue à se développer et à s'enrichir très rapidement. Les causes en sont dans les exigences techniques modernes. En effet ces derniers temps on a vu un développement intense des systèmes de commande automatique. Ainsi, les exigences imposées à la base théorique de la commande automatique deviennent ainsi de plus en plus sérieuses. Le développement de cette théorie est impossible sans l'analyse des erreurs accompagnant inévitablement les processus de commande se déroulant toujours sous l'action de perturbations aléatoires (appelées « bruits »). Ces perturbations sont de par leur nature des fonctions aléatoires. Pour choisir d'une manière rationnelle les paramètres des systèmes

de commande, il y a lieu d'étudier leur réaction à l'action continue de perturbations aléatoires, or, l'unique outil efficace et pratique est la théorie des fonctions aléatoires.

Dans le présent chapitre nous allons étudier les notions fondamentales de cette théorie et la manière dont se posent dans la pratique les problèmes exigeant l'utilisation de la théorie des fonctions aléatoires. De plus, nous exposerons ici les règles générales permettant d'utiliser les caractéristiques des fonctions aléatoires analogues à celles qui ont été formulées à propos des caractéristiques numériques des variables aléatoires ordinaires.

La première notion fondamentale que nous rencontrerons, c'est celle *de fonction aléatoire*. Cette notion est autant plus étendue et riche que celle de variable aléatoire que le sont les notions de variable et de fonction comparées à la notion de grandeur constante.

Revenons à la définition de la variable aléatoire. On appelle variable aléatoire une grandeur pouvant dans l'expérience prendre une valeur quelconque inconnue à l'avance. D'une manière analogue nous allons définir une fonction aléatoire.

Une *fonction aléatoire* est une fonction pouvant dans l'expérience adopter une forme quelconque inconnue d'avance.

La forme prise par une fonction aléatoire lors d'une expérience est appelée *réalisation* de cette fonction aléatoire. Si on effectue une série d'expériences sur une fonction aléatoire, on obtient une famille de réalisations de cette fonction.

Nous allons donner plusieurs exemples de fonctions aléatoires.

Exemple 1. Théoriquement la tension U du réseau est constante et égale à 220 V. En fait la tension oscille tout le temps autour de la valeur nominale $U_0 = 220$ V et se trouve être une fonction aléatoire du temps: $U(t)$. La mesure de cette tension durant un certain laps de temps τ est une expérience conduisant à une certaine forme (réalisation) de la fonction aléatoire $U(t)$ (voir fig. 15.1.1).

Si l'on trace automatiquement à l'aide d'un appareil enregistreur la courbe $U(t)$ et que l'on effectue une série d'expériences (dont chacune dure τ), dans chaque expérience l'appareil enregistrera une nouvelle réalisation de la fonction aléatoire $U(t)$; la série d'expériences donnera une famille de réalisations (fig. 15.1.2).

Exemple 2. La température de l'air $T(t)$ mesurée durant une journée est une fonction aléatoire du temps t. En enregistrant les valeurs de la température pour différents jours d'une même saison en fonction du temps on obtient une famille de réalisations de cette fonction aléatoire (fig. 15.1.3).

Exemple 3. Un radar au sol guide un avion jusqu'à une altitude égale à h. Le processus d'ascension est une fonction aléatoire $H(t)$. L'enregistrement répété des variations de l'altitude à l'aide d'un appareil donne une famille de réalisations de la fonction aléatoire $H(t)$ (fig. 15.1.4).

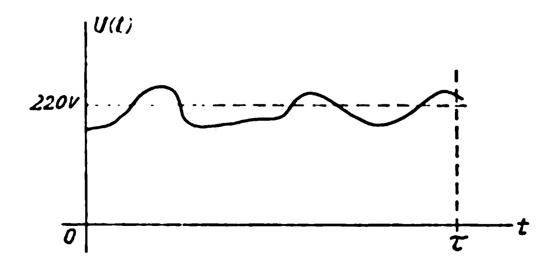


Fig. 15.1.1

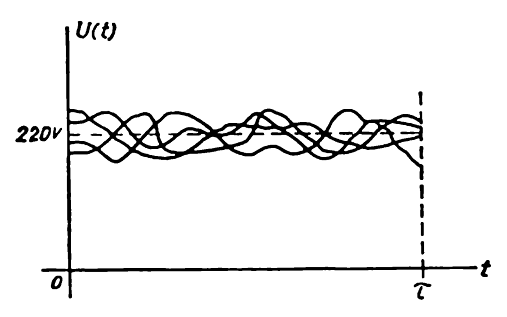


Fig. 15.1.2



Fig. 15.1.3

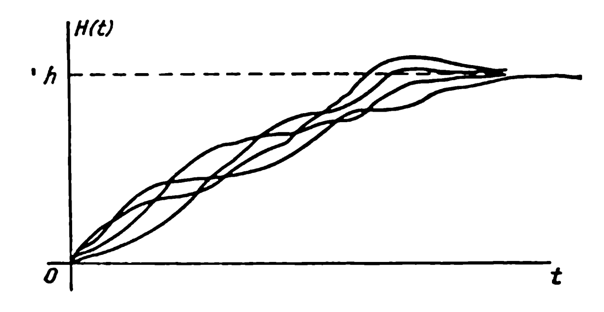


Fig. 15.1.4

La liste des exemples de fonctions aléatoires que l'on rencontre dans les applications techniques est sans fin. En effet, chaque fois que nous avons affaire à un système travaillant d'une manière continue (système de mesure, de commande, de guidage, de régulation), lors de l'analyse de sa précision de fonctionnement il y lieu de tenir compte d'actions aléatoires (bruits). Tant les bruits que les réactions produites par ceux-ci sont des fonctions aléatoires du temps.

Jusqu'à présent nous n'avons parlé que des fonctions aléatoires dont l'argument est le temps t. Dans la pratique on rencontre bien des fonctions aléatoires dépendant non pas du temps mais d'autres

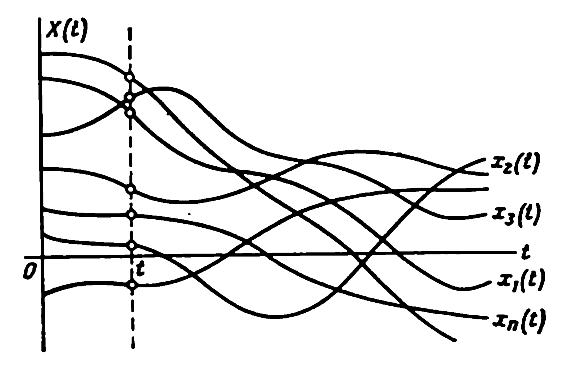


Fig. 15.1.5

arguments. Par exemple, les caractéristiques de résistance d'une barre hétérogène peuvent être considérées comme des fonctions aléatoires de l'abscisse x de la section. La température de l'air dans les différentes couches de l'atmosphère peut être considérée comme une fonction aléatoire de l'altitude H.

Dans la pratique on rencontre également des fonctions aléatoires dépendant non pas d'un seul, mais de plusieurs arguments. Par exemple, les données aérologiques caractérisant l'état de l'atmosphère (température, pression, vent) sont dans le cas général des fonctions aléatoires de quatre arguments: de trois coordonnées x, y, z et du temps t.

Dans ce cours nous n'étudierons que des fonctions aléatoires d'un seul argument. Comme le plus souvent cet argument est le temps, nous le désignerons par la lettre t. De plus, nous convenons de désigner les fonctions aléatoires par des lettres majuscules $[X(t), Y(t), \ldots]$ pour les distinguer des fonctions non aléatoires que nous désignerons par des minuscules: $[x(t), y(t), \ldots]$.

Soit une certaine fonction aléatoire $X(t)$. Supposons qu'elle ait été soumise à n expériences indépendantes ayant donné n réalisations $x_1(t), x_2(t), \ldots, x_n(t)$ (fig. 15.1.5), l'indice désignant le numéro de l'expérience.

Evidemment chaque réalisation *aléatoire* de la fonction aléatoire $X(t)$ est une fonction *non aléatoire*.

Fixons maintenant une certaine valeur de l'argument t et voyons ce que devient la fonction aléatoire $X(t)$. Il est clair qu'elle deviendra une *variable aléatoire* au sens habituel de ce terme. Nous appelons cette variable aléatoire *section* de la fonction aléatoire correspondant

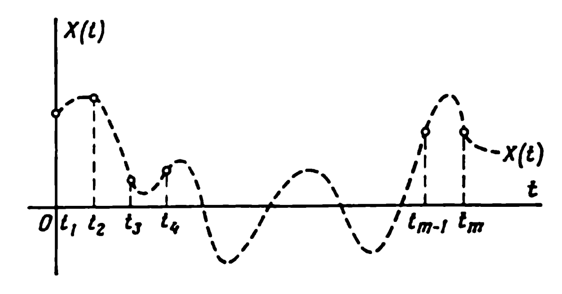


Fig. 15.2.1

à t donné. Si l'on considère la section d'une famille de réalisations à t donné (fig. 15.1.5), on obtient n valeurs prises par la variable aléatoire $X(t)$ dans n expériences.

Nous voyons qu'une fonction aléatoire possède en même temps les propriétés d'une variable aléatoire et d'une fonction. Si l'on fixe la valeur de l'argument, elle devient une variable aléatoire habituelle; chaque expérience en fournit une réalisation qui est une fonction ordinaire (non aléatoire).

Dans l'exposé ultérieur nous envisagerons une même fonction $X(t)$ tantôt comme une fonction aléatoire, tantôt comme une variable aléatoire, suivant qu'elle est considérée sur toute la gamme de variation de t ou pour une de ses valeurs.

15.2. Notion de fonction aléatoire en tant que généralisation de la notion de système de variables aléatoires. Loi de répartition d'une fonction aléatoire

Considérons une fonction aléatoire $X(t)$ sur un certain intervalle de temps (fig. 15.2.1).

En toute rigueur, on ne peut pas représenter une fonction aléatoire par un graphique; on ne peut figurer qu'une de ses réalisations. Cependant d'une manière toute conventionnelle on peut dessiner la fonction aléatoire $X(t)$ comme une courbe, celle-ci étant non pas une réalisation mais tout l'ensemble des réalisations possibles de $X(t)$. Pour noter qu'il s'agit d'une courbe conventionnelle nous la représenterons en pointillé.

Supposons que les variations d'une fonction aléatoire soient enregistrées à l'aide d'un appareil et ceci non pas d'une manière continue mais à des instants $t_1, t_2, \ldots, t_m$.

Comme nous l'avons déjà noté, pour une valeur donnée de t la fonction aléatoire devient une variable aléatoire ordinaire. Par conséquent, les résultats de l'enregistrement forment un système de m variables aléatoires:

$$X(t_1),\ X(t_2),\ \ldots,\ X(t_m). \qquad (15.2.1)$$

Il est évident que l'enregistrement d'une fonction aléatoire à des intervalles de temps suffisamment rapprochés donne une image assez précise de ses variations. Ainsi, l'étude d'une fonction aléatoire peut à une certaine approximation être ramenée à celle d'un système de variables aléatoires (15.2.1). Au fur et à mesure que m augmente, cette approximation devient de plus en plus précise. A la limite le nombre de valeurs de l'argument et par conséquent celui de variables aléatoires (15.2.1) devient infini. Aussi la notion de fonction aléatoire peut-elle être considérée comme la généralisation naturelle de la notion de système des variables aléatoires au cas d'un ensemble infini (innombrable) de variables aléatoires formant un système.

En partant de cette interprétation de la fonction aléatoire, proposons-nous de répondre à la question suivante: qu'entend-on par la *loi de répartition* d'une fonction aléatoire?

On sait que la loi de répartition d'une variable aléatoire est une fonction d'un seul argument, celle d'un système de deux variables une fonction de deux arguments, etc. Cependant, l'utilisation pratique de fonctions d'arguments multiples en qualité de caractéristiques probabilistes est si peu commode que même pour un système de trois ou quatre variables on abandonne ces lois de répartition pour ne se servir que des caractéristiques numériques. Quant à la loi de répartition de la variable aléatoire, fonction d'un nombre infini d'arguments, dans le meilleur des cas on ne peut que lui donner formellement une forme symbolique quelconque, l'utilisation pratique de cette caractéristique étant évidemment impossible.

Néanmoins, on peut trouver pour une fonction aléatoire certaines caractéristiques probabilistes analogues aux lois de répartition. Précisons notre idée.

Considérons la variable aléatoire $X(t)$ qui est la section d'une fonction aléatoire à l'instant t (fig 15.2.2). Cette variable aléatoire possède une loi de répartition qui dépend évidemment, dans le cas général, de t. Désignons-la par $f(x, t)$; cette fonction est appelée loi de répartition *unidimensionnelle* de la fonction aléatoire $X(t)$.

Evidemment, la fonction $f(x, t)$ n'est pas une caractéristique exhaustive de $X(t)$. En effet, elle ne donne que la loi de répartition de $X(t)$ pour une valeur fixée, bien qu'arbitraire de t, et ne montre

pas comment les variables aléatoires $X(t)$ varient pour différents t. La loi de répartition *bidimensionnelle*

$$f(x_1, x_2; t_1, t_2) \tag{15.2.2}$$

représente une caractéristique plus complète de la fonction aléatoire $X(t)$. C'est la loi de répartition d'un système de deux variables aléatoires $X(t_1)$, $X(t_2)$, c'est-à-dire de deux sections arbitraires de la fonction aléatoire $X(t)$. Cependant cette caractéristique n'est pas non plus complète; la loi de répartition *tridimensionnelle*

$$f(x_1, x_2, x_3; t_1, t_2, t_3) \tag{15.2.3}$$

serait une caractéristique encore plus complète.

Il est évident que théoriquement on peut augmenter indéfiniment le nombre d'arguments pour avoir des caractéristiques de plus en

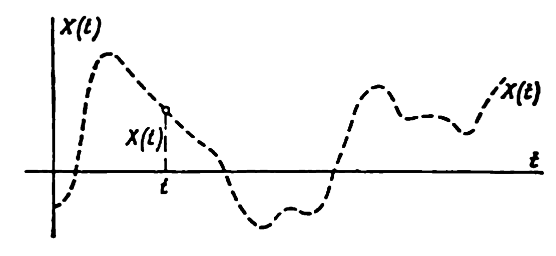


Fig. 15.2.2

plus détaillées et complètes de la fonction aléatoire, mais l'utilisation des caractéristiques aussi compliquées, dépendant d'un grand nombre d'arguments, est extrêmement difficile. C'est pourquoi dans l'étude des lois de répartition des fonctions aléatoires, on se limite généralement aux cas particuliers où, pour caractériser complètement une fonction aléatoire, il suffit, par exemple, de connaître la fonction (15.2.2) (processus sans « postaction »).

Dans notre exposé élémentaire des fonctions aléatoires nous n'utiliserons pas les lois de répartition, en nous bornant aux caractéristiques simples des fonctions aléatoires, analogues aux caractéristiques numériques des variables aléatoires.

15.3. Caractéristiques des fonctions aléatoires

Maintes fois nous avons eu l'occasion de nous rendre compte du rôle important que jouent en théorie des probabilités les caractéristiques numériques fondamentales des variables aléatoires, à savoir l'espérance mathématique et la variance dans le cas d'une seule variable aléatoire, l'espérance mathématique et la matrice de corrélation pour un système de variables aléatoires. L'art d'utiliser autant que

possible les caractéristiques numériques, sans recourir aux lois de répartition, se trouve à la base des applications de la théorie des probabilités. L'appareil des caractéristiques numériques est un outil très souple et puissant, permettant de résoudre assez simplement un grand nombre de problèmes pratiques.

Un appareil semblable est utilisé dans la théorie des fonctions aléatoires. On introduit pour les fonctions aléatoires des caractéristiques de base simples, analogues aux caractéristiques numériques des variables aléatoires, et l'on en établit le mode d'emploi. Il se trouve que cet appareil est suffisant dans de nombreux problèmes pratiques.

A la différence des caractéristiques numériques des variables aléatoires qui sont certains *nombres*, les caractéristiques des fonctions aléatoires sont dans le cas général des *fonctions*.

On définit comme suit l'espérance mathématique de la fonction aléatoire $X(t)$. Soit la section de la fonction aléatoire $X(t)$ pour t donné. Dans cette section nous avons une variable aléatoire habituelle et nous allons en déterminer l'espérance mathématique. Il est évident que dans le cas général elle dépend de t:

$$m_x(t) = M[X(t)]. \qquad (15.3.1)$$

Ainsi, l'*espérance mathématique d'une fonction aléatoire* $X(t)$ *est une fonction non aléatoire* $m_x(t)$ *qui, pour toute valeur de l'argument* t, *est égale à l'espérance mathématique de la section correspondante de la fonction aléatoire.*

L'espérance mathématique d'une fonction aléatoire est une certaine *fonction moyenne* autour de laquelle varient les diverses réalisations de cette fonction aléatoire.

Sur la figure 15.3.1 on a représenté en trait fin les réalisations d'une fonction aléatoire et en trait gras son espérance mathématique.

On définit d'une manière analogue la variance d'une fonction aléatoire.

On appelle variance d'une fonction aléatoire $X(t)$ *une fonction non aléatoire* $D_x(t)$ *dont la valeur pour* t *quelconque est égale à la variance de la section correspondante de la fonction aléatoire*:

$$D_x(t) = D[X(t)]. \qquad (15.3.2)$$

La variance d'une fonction aléatoire caractérise, pour t fixé, la dispersion des réalisations possibles de la fonction aléatoire par rapport à la moyenne, en d'autres termes, montre l'influence du hasard sur la fonction aléatoire.

Il est évident que $D_x(t)$ est une fonction non négative. La racine carrée de cette fonction, $\sigma_x(t)$, est l'écart quadratique moyen de la fonction aléatoire, soit:

$$\sigma_x(t) = \sqrt{D_x(t)}. \qquad (15.3.3)$$

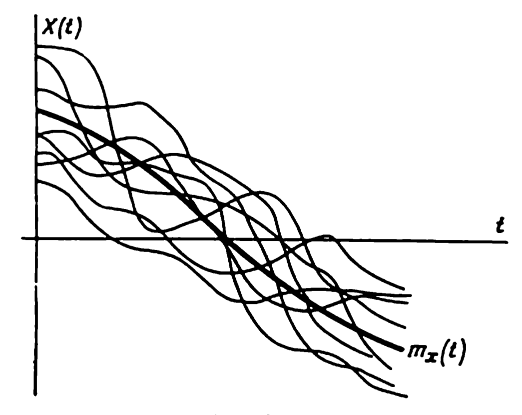


Fig. 15.3.1

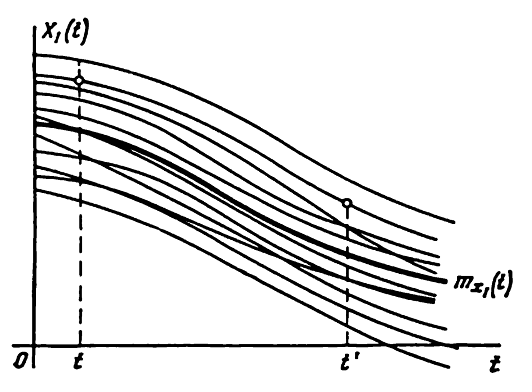


Fig. 15.3.2

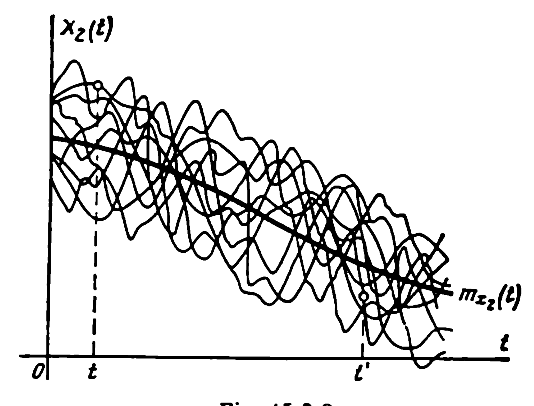


Fig. 15.3.3

L'espérance mathématique et la variance sont des caractéristiques très importantes d'une fonction aléatoire, cependant elles sont insuffisantes pour en décrire les particularités essentielles. Pour s'en rendre compte, considérons deux fonctions aléatoires $X_1(t)$ et $X_2(t)$ représentées à titre d'exemple par des familles de réalisations (fig. 15.3.2 et 15.3.3).

Les fonctions aléatoires $X_1(t)$ et $X_2(t)$ ont des espérances mathématiques et des variances à peu près identiques, cependant elles sont très différentes. La fonction $X_1(t)$ (fig. 15.3.2) varie d'une manière continue. Si, par exemple, au point t sa valeur dépasse nettement la moyenne, il est très probable qu'au point t' elle aura

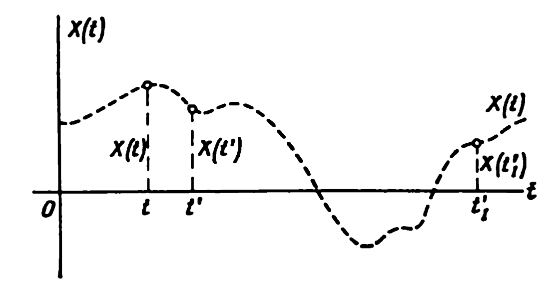


Fig. 15.3.4

également une valeur supérieure à la moyenne. Les valeurs de $X_1(t)$ dépendent les unes des autres pour différents t. Au contraire, la fonction aléatoire $X_2(t)$ (fig. 15.3.3) a un caractère oscillatoire très net, les oscillations étant irrégulières et désordonnées. Une telle fonction aléatoire est caractérisée par un affaiblissement des liens entre ses valeurs au fur et à mesure qu'augmente la distance les séparant suivant l'axe des t.

Il est évident que la structure de ces deux processus aléatoires est tout à fait différente, mais cette différence n'apparaît ni dans l'espérance mathématique ni dans la variance; ainsi, il y a lieu d'introduire une caractéristique spéciale pour décrire une fonction aléatoire. Cette caractéristique est appelée *fonction de corrélation* (ou *fonction d'autocorrélation*). La fonction de corrélation caractérise le degré de dépendance entre les sections d'une fonction aléatoire pour différentes valeurs de t.

Soit la fonction aléatoire $X(t)$ (fig. 15.3.4); considérons deux de ses sections pour deux instants différents t et t', c'est-à-dire les deux variables aléatoires $X(t)$ et $X(t')$. Il est évident que pour des instants voisins t et t' les variables $X(t)$ et $X(t')$ sont étroitement liées entre elles; si la variable $X(t)$ a pris une certaine valeur, $X(t')$ prendra avec une grande probabilité une valeur voisine. Il est également évi-

dent qu'au fur et à mesure de l'augmentation de l'intervalle entre les sections t et t' le lien entre $X(t)$ et $X(t')$ doit en général s'affaiblir.

Le degré d'interdépendance des variables $X(t)$ et $X(t')$ peut être, en une grande mesure, caractérisé par leur *covariance*; naturellement cette dernière est une fonction des deux arguments t et t'. Cette fonction est appelée *fonction de corrélation.*

Ainsi, on appelle *fonction de corrélation de la fonction aléatoire* $X(t)$ *une fonction non aléatoire de deux arguments* $K_x(t, t')$ *qui pour tout couple de valeurs* t, t' *est égale à la covariance des sections correspondantes de la fonction aléatoire*:

$$K_x(t, t') = M[\overset{\circ}{X}(t)\overset{\circ}{X}(t')], \qquad (15.3.4)$$

où

$$\overset{\circ}{X}(t) = X(t) - m_x(t), \quad \overset{\circ}{X}(t') = X(t') - m_x(t').$$

Revenons aux fonctions aléatoires $X_1(t)$ et $X_2(t)$ (fig. 15.3.2 et 15.3.3). Nous voyons que bien que leurs espérances mathématiques et variances soient à peu près les mêmes, leurs fonctions de corrélation sont très différentes. Celle de $X_1(t)$ décroît lentement avec l'augmentation de l'intervalle (t, t'); au contraire la fonction de corrélation de $X_2(t)$ décroît rapidement au fur et à mesure qu'augmente cet intervalle.

Voyons ce que devient la fonction de corrélation $K_x(t, t')$ lorsque ses arguments se confondent. Posant $t = t'$, on a:

$$K_x(t, t) = M[(\overset{\circ}{X}(t))^2] = D_x(t), \qquad (15.3.5)$$

c'est-à-dire que pour $t' = t$ la *fonction de corrélation devient la variance de la fonction aléatoire.*

Ainsi, la variance d'une fonction aléatoire en tant que sa caractéristique n'est d'aucune utilité; les caractéristiques fondamentales d'une fonction aléatoire sont donc l'espérance mathématique et la fonction de corrélation.

La covariance de deux variables aléatoires $X(t)$ et $X(t')$ ne dépendant pas de leur ordre de présentation, la fonction de corrélation est symétrique par rapport à ses arguments, c'est-à-dire qu'elle ne change pas lorsqu'on intervertit les arguments:

$$K_x(t, t') = K_x(t', t). \qquad (15.3.6)$$

Si l'on représente la fonction de corrélation $K_x(t, t')$ sous la forme d'une surface, celle-ci sera symétrique par rapport au plan vertical Q passant par la bissectrice de l'angle tOt' (fig. 15.3.5).

Il est à noter que les propriétés des fonctions de corrélation découlent naturellement de celles de la matrice de corrélation du système de variables aléatoires. En effet, remplaçons approximativement la

fonction aléatoire $X(t)$ par un système de m variables aléatoires $X(t_1), X(t_2), \ldots, X(t_m)$. Avec l'augmentation de m et la diminution respective des intervalles entre les arguments, la matrice de corrélation du système, représentant une table à deux entrées, devient à la limite une fonction de deux arguments continûment variables ayant des propriétés analogues. La symétrie de la matrice de corrélation par rapport à la diagonale principale entraîne la symétrie de la fonction de corrélation (15.3.6).

Dans la diagonale principale de la matrice de corrélation on trouve les variances des variables aléatoires; d'une manière analogue pour $t = t'$ la fonction de corrélation $K_x(t, t')$ devient la variance $D_x(t)$.

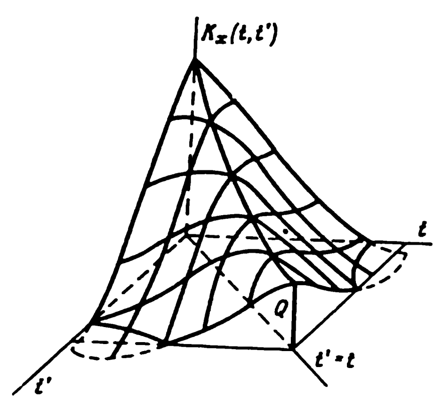


Fig. 15.3.5

En pratique lorsqu'on a besoin de construire la fonction de corrélation de la fonction aléatoire $X(t)$, on procède généralement comme suit: on se donne plusieurs valeurs d'argument équidistantes et l'on détermine la matrice de corrélation pour ce système de variables aléatoires. Cette matrice n'est rien d'autre que la table des valeurs de la fonction de corrélation pour un réseau rectangulaire de valeurs des arguments dans le plan (t, t'). Puis par interpolation ou par approximation on peut déterminer la fonction de deux arguments $K_x(t, t')$.

A la place de la fonction de corrélation $K_x(t, t')$ on peut utiliser la *fonction de corrélation normée*:

$$r_x(t, t') = \frac{K_x(t, t')}{\sigma_x(t)\,\sigma_x(t')}, \tag{15.3.7}$$

qui est le coefficient de corrélation des variables $X(t)$, $X(t')$. La fonction de corrélation normée est analogue à la matrice de corrélation normée d'un système de variables aléatoires. Pour $t = t'$ la fonction de corrélation normée est égale à l'unité:

$$r_x(t, t) = \frac{K_x(t, t)}{[\sigma_x(t)]^2} = \frac{D_x(t)}{[\sigma_x(t)]^2} = 1. \tag{15.3.8}$$

Examinons quel effet produisent sur les caractéristiques fondamentales d'une fonction aléatoire des opérations élémentaires, à savoir, l'addition d'un terme non aléatoire et la multiplication par un facteur non aléatoire. Ces termes et facteurs non aléatoires peuvent être tant des grandeurs constantes que, dans le cas général, des fonctions de t.

Ajoutons à la fonction aléatoire $X(t)$ le terme non aléatoire $\varphi(t)$. On trouve ainsi la nouvelle fonction aléatoire:

$$Y(t) = X(t) + \varphi(t). \tag{15.3.9}$$

En vertu du théorème d'addition des espérances mathématiques:

$$m_y(t) = m_x(t) + \varphi(t), \tag{15.3.10}$$

ainsi *lorsqu'on ajoute à une fonction aléatoire un terme non aléatoire, ce même terme non aléatoire doit être ajouté à l'espérance mathématique de la fonction aléatoire.*

Calculons la fonction de corrélation de la fonction aléatoire $Y(t)$:

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')] = M[(Y(t) - m_y(t))(Y(t') - m_y(t')] = \\ = M[(X(t) + \varphi(t) - m_x(t) - \varphi(t))(X(t') + \varphi(t') - \\ - m_x(t') - \varphi(t'))] = M[(X(t) - m_x(t))(X(t') - m_x(t'))] = K_x(t, t'). \tag{15.3.11}$$

Ainsi, *lorsqu'on ajoute un terme certain à une fonction aléatoire, la fonction de corrélation de cette fonction aléatoire ne s'en trouve pas changée.*

Multiplions la fonction aléatoire $X(t)$ par une fonction certaine $\varphi(t)$:

$$Y(t) = \varphi(t) X(t). \tag{15.3.12}$$

En faisant sortir la grandeur non aléatoire $\varphi(t)$ du signe de l'espérance mathématique on a:

$$m_y(t) = M[\varphi(t) X(t)] = \varphi(t) m_x(t), \tag{15.3.13}$$

c'est-à-dire que *si on multiplie une fonction aléatoire par un facteur certain, son espérance mathématique se trouve multipliée par ce même facteur.*

Calculons la fonction de corrélation:

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')] = M[(Y(t) - m_y(t))(Y(t') - \\ - m_y(t'))] = M[\varphi(t)\varphi(t')(X(t) - m_x(t))(X(t') - \\ - m_x(t'))] = \varphi(t)\varphi(t') K_x(t, t'), \tag{15.3.14}$$

c'est-à-dire que *lorsqu'on multiplie une fonction aléatoire par une fonction certaine $\varphi(t)$, sa fonction de corrélation se trouve multipliée par $\varphi(t)\varphi(t')$.*

En particulier, lorsque $\varphi(t) = c$ (const.), la fonction de corrélation se trouve multipliée par c^2.

En utilisant les propriétés décrites ci-dessus des caractéristiques des fonctions aléatoires, on peut dans un certain nombre de cas simplifier beaucoup les calculs. En particulier, lorsqu'il y a lieu d'étu-

dier la fonction de corrélation ou la variance d'une fonction aléatoire, il sera légitime d'envisager la *fonction centrée*:

$$\overset{\circ}{X}(t) = X(t) - m_x(t). \tag{15.3.15}$$

L'espérance mathématique d'une fonction centrée est identiquement nulle, et sa fonction de corrélation est celle de la fonction aléatoire $X(t)$:

$$K_{\overset{\circ}{x}}(t, t') = M[\overset{\circ}{X}(t)\overset{\circ}{X}(t')] = K_x(t, t'). \tag{15.3.16}$$

Dans la suite lorsque nous étudierons des questions liées aux propriétés corrélationnelles des fonctions aléatoires, nous passerons toujours aux fonctions centrées correspondantes que nous noterons du signe ° placé au-dessus du symbole de la fonction.

Parfois en plus du centrage des fonctions aléatoires on recourt à la normalisation de ces dernières. On appelle fonction *normée* une fonction de la forme:

$$X_N(t) = \frac{\overset{\circ}{X}(t)}{\sigma_x(t)}. \tag{15.3.17}$$

La fonction de corrélation de la fonction aléatoire normée $X_N(t)$ est:

$$K_{x_N}(t, t') = \frac{K_x(t, t')}{\sigma_x(t)\,\sigma_x(t')} = r_x(t, t'), \tag{15.3.18}$$

et sa variance est égale à l'unité.

15.4. Détermination expérimentale des caractéristiques d'une fonction aléatoire

Supposons que n expériences (observations) indépendantes de la fonction aléatoire $X(t)$ aient donné n réalisations de cette fonction (fig. 15.4.1).

Trouver les estimations des caractéristiques suivantes de la fonction aléatoire: de son espérance mathématique $m_x(t)$, de sa variance $D_x(t)$ et de sa fonction de corrélation $K_x(t, t')$.

Considérons à cet effet plusieurs sections de la fonction aléatoire aux instants

$$t_1, t_2, \ldots, t_m$$

et enregistrons les n valeurs prises par la fonction $X(t)$ à chaque instant.

Généralement les sections $t_1, t_2, \ldots, t_m$ sont équidistantes; la longueur de l'intervalle les séparant dépend de la forme des courbes expérimentales et se choisit de telle sorte qu'il soit possible, d'après les points obtenus, de restituer l'allure générale des courbes. Souvent,

l'intervalle entre des valeurs voisines de t est donné, indépendamment du problème traité, par la fréquence de fonctionnement de l'appareil enregistreur (par exemple, par la vitesse de la caméra employée).

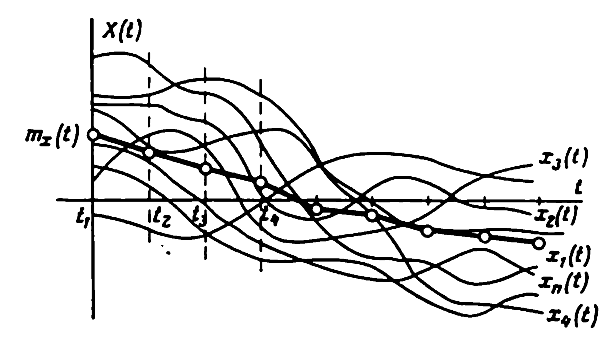


Fig. 15.4.1

Les valeurs enregistrées $X(t)$ sont portées dans un tableau dont chaque ligne correspond à une certaine réalisation, le nombre de colonnes étant égal au nombre de « repères » de l'argument (tableau 15.4.1).

Tableau 15.4.1

$X(t)$ \ t	t_1	t_2	...	t_k	...	t_l	...	t_m
$x_1(t)$	$x_1(t_1)$	$x_1(t_2)$	...	$x_1(t_k)$	...	$x_1(t_l)$	...	$x_1(t_m)$
$x_2(t)$	$x_2(t_1)$	$x_2(t_2)$	...	$x_2(t_k)$	...	$x_2(t_l)$	...	$x_2(t_m)$
...	...	...	...	...	...	...	...	...
$x_i(t)$	$x_i(t_1)$	$x_i(t_2)$	...	$x_i(t_k)$	...	$x_i(t_l)$	...	$x_i(t_m)$
...	...	...	...	...	...	...	...	...
$x_n(t)$	$x_n(t_1)$	$x_n(t_2)$	...	$x_n(t_k)$	...	$x_n(t_l)$	...	$x_n(t_m)$

Dans l'i-ème ligne du tableau 15.4.1 on trouve les valeurs de la fonction aléatoire observées dans l'i-ème réalisation (l'i-ème expérience) pour des valeurs de l'argument $t_1, t_2, \ldots, t_m$. Le symbole $x_i(t_k)$ désigne la valeur correspondant à l'i-ème réalisation à l'instant t_k.

Les données obtenues ne sont rien d'autre que les résultats de n expériences sur un système de m variables aléatoires

$$X(t_1),\ X(t_2),\ \ldots,\ X(t_m)$$

et doivent être traitées comme tels (voir § 14.3). Tout d'abord on trouve les estimations des espérances mathématiques:

$$\tilde{m}_x(t_k)=\frac{\sum_{i=1}^{n} x_i(t_k)}{n}, \qquad (15.4.1)$$

puis celles de la variance:

$$\tilde{D}_x(t_k)=\frac{\sum_{i=1}^{n}[x_i(t_k)-\tilde{m}_x(t_k)]^2}{n-1} \qquad (15.4.2)$$

et enfin les estimations des covariances

$$\tilde{K}_x(t_k, t_l)=\frac{\sum_{i=1}^{n}[x_i(t_k)-\tilde{m}_x(t_k)][x_i(t_l)-\tilde{m}_x(t_l)]}{n-1}. \qquad (15.4.3)$$

Souvent il est commode, lors du calcul des estimations des variances et des covariances d'utiliser la relation existant entre les moments initiaux et centrés; les formules respectives sont:

$$\tilde{D}_x(t_k)=\left[\frac{\sum_{i=1}^{n}[x_i(t_k)]^2}{n}-[\tilde{m}_x(t_k)]^2\right]\frac{n}{n-1}; \qquad (15.4.4)$$

$$\tilde{K}_x(t_k, t_l)=\left[\frac{\sum_{i=1}^{n}x_i(t_k)\,x_i(t_l)}{n}-\tilde{m}_x(t_k)\tilde{m}_x(t_l)\right]\frac{n}{n-1}. \qquad (15.4.5)$$

Lorsqu'on utilise ces dernières formules, afin d'éviter d'avoir affaire à des différences de nombres voisins, on transfère l'origine des ordonnées plus près de l'espérance mathématique.

Après avoir calculé ces caractéristiques, on peut en utilisant la suite des valeurs $\tilde{m}_x(t_1), \tilde{m}_x(t_2), \ldots, \tilde{m}_x(t_m)$ déterminer la fonction $\tilde{m}_x(t)$ (fig. 15.4.1). On construit d'une manière analogue la courbe $\tilde{D}_x(t)$. La fonction des deux arguments $\tilde{K}_x(t, t')$ s'obtient d'après ses valeurs dans un réseau rectangulaire de points. En cas de nécessité toutes ces fonctions peuvent être approchées par des expressions analytiques.

15.5. Méthodes de calcul des caractéristiques des fonctions aléatoires transformées d'après celles de fonctions initiales

Dans le paragraphe précédent nous avons montré comment on calcule les caractéristiques d'une fonction aléatoire à partir des données expérimentales. Cette méthode n'est pas toujours utilisable. Tout

d'abord les expériences spéciales nécessaires à la détermination des fonctions aléatoires qui nous intéressent peuvent être coûteuses et compliquées. Deuxièmement, souvent il y a lieu d'étudier des fonctions aléatoires caractérisant les erreurs des appareils de mesure, des dispositifs de mire, des systèmes de commande, etc. en cours de conception et n'existant pas encore. On étudie alors ces erreurs éventuelles justement dans le but de choisir les paramètres du système de telle sorte que les erreurs soient minimales. Il est évident que dans ce cas l'étude directe des fonctions aléatoires caractérisant le fonctionnement du système est inutile sinon même impossible. Dans ces cas on fait alors appel à des méthodes *indirectes*. Nous avons déjà utilisé des méthodes indirectes dans les chapitres 10, 11, 12 de notre cours, où nous avons montré comment on recherche les lois de répartition et les caractéristiques numériques des variables aléatoires d'après celles d'autres variables aléatoires, liées aux premières. En utilisant des méthodes analogues on peut déterminer indirectement les caractéristiques des fonctions aléatoires d'après les caractéristiques d'autres fonctions aléatoires, liées aux premières. La théorie appliquée des fonctions aléatoires est essentiellement consacrée au développement de ces méthodes indirectes.

Dans la pratique le problème de l'étude indirecte des fonctions aléatoires se pose de la manière suivante.

Considérons un système dynamique quelconque A ; par « système dynamique » on entend un appareil quelconque, un viseur, un calculateur, un système de commande automatique, etc. Ce système peut être mécanique, électrique ou contenir des éléments de toute autre nature. On peut se représenter comme suit le fonctionnement du système : l'entrée du système reçoit d'une manière continue certaines données ; le système en effectue le traitement et délivre le résultat également d'une manière continue. Convenons d'appeler « action » les données arrivant à l'entrée du système et « réponse » le résultat fourni par le système comme réponse à l'action. Les actions peuvent être des tensions variables, des coordonnées angulaires et linéaires des objectifs, des signaux ou des commandes reçus par un système de commande, etc. De même, la réponse du système peut être élaborée sous différentes formes : tensions, déplacements angulaires, etc.

Considérons le cas le plus simple, lorsque l'entrée du système A reçoit une seule action, une certaine fonction du temps $x(t)$; la réponse à cette action est une autre fonction du temps $y(t)$ (fig. 15.5.1).

Nous dirons que le système A opère sur l'action d'entrée une certaine *transformation* $x(t) \to y(t)$ que nous écrirons comme suit :

$$y(t) = A\{x(t)\}. \tag{15.5.1}$$

La transformation A peut être plus ou moins compliquée et prendre une forme quelconque. Dans les cas les plus simples c'est la multiplication par un facteur donné (amplificateurs, mécanismes de mul-

tiplication), la dérivation et l'intégration (dispositifs dérivateurs et intégrateurs). Cependant on rencontre rarement dans la pratique des systèmes réalisant des transformations simples; en général le fonctionnement d'un système est décrit par des équations différentielles, et la transformation A revient à la résolution d'une équation différentielle reliant l'action $x(t)$ et la réponse $y(t)$.

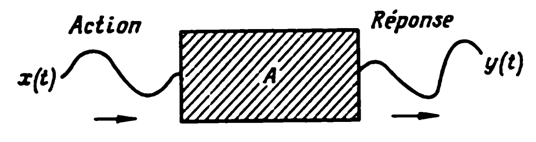


Fig. 15.5.1

Lors de l'étude d'un système dynamique on résout tout d'abord le problème essentiel qui consiste à trouver la réponse $y(t)$ du système à uneaction donnée $x(t)$. Cependant ceci ne permet pas de faire une étude complète du système et d'estimer sa qualité technique. En

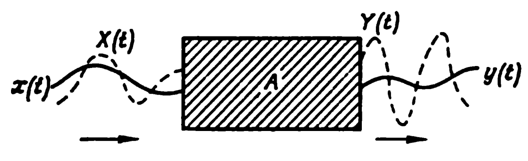


Fig. 15.5.2

réalité l'action $x(t)$ n'est jamais un signal pur, elle est toujours affectée par des erreurs aléatoires (perturbations) qui font que le système reçoit à l'entrée non pas la fonction donnée $x(t)$ mais une fonction aléatoire $X(t)$; par conséquent sa réponse est également une fonction aléatoire $Y(t)$ différente évidemment de la réponse théorique $y(t)$ (fig. 15.5.2).

Une question se pose tout naturellement: à savoir quelle est la grandeur des déformations aléatoires altérant la réponse du système en présence de perturbations aléatoires agissant à l'entrée? Et comment faut-il choisir les paramètres du système pour que ces déformations soient minimales?

Ces problèmes ne peuvent pas être résolus par les méthodes classiques de la théorie des probabilités; seule la théorie des fonctions aléatoires permet de les aborder.

Le premier des deux problèmes posés ci-dessus, le problème direct, est évidemment le plus simple. Nous allons formuler ce problème de la manière suivante.

L'entrée du système dynamique A est attaquée par une fonction aléatoire $X(t)$; le système effectue une certaine transformation

donnant à la sortie la fonction aléatoire

$$Y(t) = A\{X(t)\}. \tag{15.5.2}$$

On connaît l'espérance mathématique et la fonction de corrélation de la fonction aléatoire $X(t)$. Il y a lieu de trouver les caractéristiques analogues de la fonction aléatoire $Y(t)$. Ou plus brièvement : trouver les caractéristiques d'une fonction aléatoire à la sortie d'un système dynamique connaissant les caractéristiques de la fonction d'entrée.

Le problème posé peut être résolu exactement dans un cas particulier, important pour la pratique, lorsque la transformation A appartient à la classe des *transformations linéaires*, on dit alors que le système A appartient à la classe des *systèmes linéaires*.

Ces notions seront étudiées dans le paragraphe suivant.

15.6. Opérateurs linéaires et non linéaires. Opérateur d'un système dynamique

Dans l'exposé de la théorie des transformations des fonctions aléatoires nous allons employer la notion d'*opérateur*, couramment utilisée en mathématiques et en technique.

La notion d'opérateur est une généralisation de celle de fonction. Lorsqu'on établit une relation fonctionnelle entre les variables y et x et que l'on écrit :

$$y = f(x) \tag{15.6.1}$$

le symbole f indique la règle suivant laquelle à une valeur donnée x on fait correspondre une valeur déterminée y. Le signe f est le symbole d'une certaine *transformation* qu'il y a lieu d'imposer à x pour obtenir y. Suivant le genre de la transformation les fonctions correspondantes peuvent être linéaires ou non linéaires, algébriques, transcendantes, etc.

Des notions analogues et des symboles correspondants sont utilisés en mathématiques pour représenter une transformation d'une fonction.

Etablissons pour une certaine fonction $x(t)$ une loi de transformation en une autre fonction $y(t)$. Ecrivons cette transformation sous la forme :

$$y(t) = A\{x(t)\}. \tag{15.6.2}$$

A titre d'exemple on peut citer la dérivation :

$$y(t) = \frac{dx(t)}{dt} \tag{15.6.3}$$

ou l'intégration :

$$y(t) = \int_0^t x(\tau)\, d\tau, \qquad (15.6.4)$$

etc.

La règle permettant de passer de la fonction $x(t)$ à la fonction $y(t)$ est appelée *opérateur*; par exemple on parlera d'opérateurs de dérivation, d'intégration, de résolution d'une équation différentielle, etc.

Dans notre définition de l'opérateur nous n'avons envisagé que la transformation d'une fonction $x(t)$ en une autre fonction y du même argument t. Pourtant cette conservation de l'argument n'est pas du tout obligatoire: un opérateur peut transformer la fonction $x(t)$ en une fonction $y(s)$ d'un autre argument, par exemple:

$$y(s) = \int_a^b \varphi(t, s)\, x(t)\, dt, \qquad (15.6.5)$$

où $\varphi(t, s)$ est un certaine fonction, dépendant en plus de l'argument t, d'un autre argument s.

Mais comme lors de l'analyse des erreurs des systèmes dynamiques, l'argument le plus naturel est le temps t, nous nous limiterons à l'étude des opérateurs transformant une fonction de t en une autre fonction du même argument.

Si un système dynamique transforme la fonction $x(t)$ appliquée à son entrée en une autre fonction $y(t)$:

$$y(t) = A\{x(t)\},$$

l'opérateur A est appelé *opérateur du système dynamique*.

Dans le cas général l'entrée du système est attaquée non pas par une seule mais par plusieurs fonctions; de même à sa sortie on peut voir apparaître plusieurs fonctions; dans ce cas l'opérateur du système transforme un ensemble de fonctions en un autre ensemble de fonctions. Cependant pour simplifier l'exposé nous n'envisagerons ici que le cas le plus élémentaire de transformation d'une fonction isolée en une autre.

Les transformations ou les opérateurs appliqués aux fonctions peuvent être de types différents. La classe la plus importante pour la pratique est celle des *opérateurs linéaires*.

Un opérateur L est appelé *linéaire homogène* s'il est doué des propriétés suivantes:

1) il peut être appliqué séparément à chacun des termes d'une somme:

$$L\{x_1(t) + x_2(t)\} = L\{x_1(t)\} + L\{x_2(t)\}; \qquad (15.6.6)$$

2) un facteur constant c peut être mis en facteur devant le signe de l'opérateur:

$$L\{cx(t)\} = cL\{x(t)\}. \qquad (15.6.7)$$

Il découle de la seconde propriété que pour un opérateur linéaire homogène on a:

$$L\{0\} = 0, \qquad (15.6.8)$$

c'est-à-dire que la réponse à zéro est égale à zéro.

A titre d'exemple d'opérateur linéaire homogène on peut citer:

1) l'opérateur de dérivation:

$$y(t) = \frac{dx(t)}{dt};$$

2) l'opérateur d'intégration:

$$y(t) = \int_0^t x(\tau)\,d\tau;$$

3) l'opérateur de multiplication par une fonction $\varphi(t)$:

$$y(t) = \varphi(t)\,x(t);$$

4) l'opérateur d'intégration pondérée par la fonction $\varphi(t)$:

$$y(t) = \int_0^t x(\tau)\,\varphi(\tau)\,d\tau,$$

etc.

A part les opérateurs linéaires homogènes il existe encore des opérateurs linéaires *inhomogènes*.

L'opérateur L est dit *linéaire inhomogène* s'il est formé par un opérateur linéaire homogène et par une fonction donnée $\varphi(t)$ additive:

$$L\{x(t)\} = L_0\{x(t)\} + \varphi(t), \qquad (15.6.9)$$

où L_0 est un opérateur linéaire homogène.

Donnons quelques exemples d'opérateurs linéaires inhomogènes:

1) $y(t) = \frac{dx(t)}{dt} + \varphi(t)$,

2) $y(t) = \int_0^t x(\tau)\,\varphi(\tau)\,d\tau + \varphi_1(t)$,

3) $y(t) = \varphi_1(t)\,x(t) + \varphi_2(t)$,

où $\varphi(t)$, $\varphi_1(t)$, $\varphi_2(t)$ sont des fonctions données, et $x(t)$ la fonction transformée par l'opérateur.

En mathématiques et en technique on utilise souvent une forme symbolique spéciale pour désigner les opérateurs. Dans certains cas des notations symboliques simplifient considérablement des transfor-

mations et permettent d'écrire des formules sous une forme simple et commode.

Par exemple, l'opérateur de dérivation est souvent désigné par la lettre p :

$$p = \frac{d}{dt}$$

mis en facteur devant l'expression à dériver. Ainsi les formes :

$$y(t) = px(t)$$

et

$$y(t) = \frac{dx(t)}{dt}$$

sont équivalentes.

Une double dérivation est désignée par p^2 :

$$p^2 x(t) = \frac{d^2 x(t)}{dt^2},$$

etc.

Cette forme symbolique est surtout commode pour l'écriture des équations différentielles.

Supposons par exemple que le fonctionnement du système dynamique A soit décrit par une équation différentielle linéaire à coefficients constants liant la réponse du système $y(t)$ et l'action $x(t)$. Sous la forme habituelle cette équation différentielle s'écrit :

$$a_n \frac{d^n y(t)}{dt^n} + a_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \ldots + a_1 \frac{dy(t)}{dt} + a_0 y(t) =$$
$$= b_m \frac{d^m x(t)}{dt^m} + b_{m-1} \frac{d^{m-1} x(t)}{dt^{m-1}} + \ldots + b_1 \frac{dx(t)}{dt} + b_0 x(t). \quad (15.6.10)$$

Sous la forme symbolique on a:

$$(a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0)\, y(t) =$$
$$= (b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0)\, x(t),$$

où $p = \frac{d}{dt}$ est l'opérateur de dérivation.

Pour plus de simplicité nous allons désigner comme suit les polynômes en p se trouvant dans les premier et second membres :

$$A_n(p) = a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0,$$
$$B_m(p) = b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0,$$

ce qui nous permet d'écrire l'équation en question sous une forme encore plus simple, soit :

$$A_n(p)\, y(t) = B_m(p)\, x(t). \quad (15.6.11)$$

Enfin en résolvant formellement l'équation (15.6.11) par rapport à $y(t)$, on obtient la solution de l'équation différentielle « explicite »,

soit :

$$y(t)=\frac{B_m(p)}{A_n(p)}\,x(t). \tag{15.6.12}$$

En utilisant des notations symboliques analogues on peut écrire sous forme opérationnelle une équation différentielle linéaire à coefficients variables. Habituellement cette équation s'écrit sous la forme :

$$a_n(t)\frac{d^ny(t)}{dt^n}+a_{n-1}(t)\frac{d^{n-1}y(t)}{dt^{n-1}}+\ldots+a_1(t)\frac{dy(t)}{dt}+a_0(t)\,y(t)=$$

$$=b_m(t)\frac{d^mx(t)}{dt^m}+b_{m-1}(t)\frac{d^{m-1}x(t)}{dt^{m-1}}+\ldots+b_1(t)\frac{dx(t)}{dt}+b_0(t)\,x(t). \tag{15.6.13}$$

Introduisons pour les polynômes en p à coefficients dépendant du temps t les désignations suivantes :

$$A_n(p,t)=a_n(t)\,p^n+a_{n-1}(t)\,p^{n-1}+\ldots+a_1(t)\,p+a_0(t),$$

$$B_m(p,t)=b_m(t)\,p^m+b_{m-1}(t)\,p^{m-1}+\ldots+b_1(t)\,p+b_0(t).$$

Alors l'opérateur de l'équation différentielle s'écrit :

$$y(t)=\frac{B_m(p,t)}{A_n(p,t)}\,x(t). \tag{15.6.14}$$

Nous utiliserons ultérieurement quand il sera besoin cette désignation pour les opérateurs.

Les systèmes dynamiques que l'on rencontre souvent en pratique sont dans la majorité des cas décrits par des équations différentielles linéaires. Il est facile de voir que dans ce cas l'opérateur du système est également linéaire.

Un système dynamique dont l'opérateur est linéaire est appelé *système dynamique linéaire.*

En plus des opérateurs et des systèmes linéaires on rencontre des opérateurs et des systèmes *non linéaires*. A titre d'exemple d'opérateurs non linéaires on peut citer

$$y(t)=x^2(t)\,;\;\; y(t)=\int_0^t x^3(\tau)\,d\tau\,;\;\; y(t)=\sin x(t)$$

ou bien la solution d'une équation différentielle non linéaire, par exemple :

$$y'(t)+\alpha\cos y(t)=x(t).$$

Un système dynamique dont l'opérateur n'est pas linéaire est appelé *système non linéaire*.

On rencontre souvent dans la pratique des systèmes linéaires. Cette propriété de linéarité permet d'appliquer à l'étude des erreurs des systèmes la théorie des fonctions aléatoires. Tout comme les

caractéristiques numériques des fonctions linéaires de variables aléatoires ordinaires peuvent être obtenues à partir des caractéristiques numériques des arguments, celles d'une fonction aléatoire à la sortie d'un système dynamique linéaire s'obtiennent si l'on connaît l'opérateur du système et des caractéristiques de la fonction aléatoire d'entrée.

Encore plus souvent dans la pratique on rencontre des systèmes qui ne sont pas rigoureusement linéaires mais admettant une linéarisation. Si les perturbations aléatoires à l'entrée du système sont suffisamment petites, tout système peut en fait, dans les limites de ces perturbations, être considéré comme approximativement linéaire, tout comme pour des variations aléatoires suffisamment petites des arguments presque toute fonction peut être linéarisée.

La méthode de linéarisation approchée des équations différentielles est couramment utilisée dans la théorie des erreurs des systèmes dynamiques.

Dans la suite nous n'étudierons que des systèmes dynamiques linéaires (ou linéarisables) et des opérateurs linéaires correspondants.

15.7. Transformations linéaires des fonctions aléatoires

Supposons qu'à l'entrée d'un système linéaire d'opérateur L on applique une fonction aléatoire $X(t)$ d'espérance mathématique $m_x(t)$ et de fonction de corrélation $K_x(t, t')$ données. La réponse du système est une fonction aléatoire

$$Y(t) = L\{X(t)\}. \qquad (15.7.1)$$

Il s'agit de trouver les caractéristiques de la fonction aléatoire $Y(t)$ à la sortie du système, à savoir $m_y(t)$ et $K_y(t, t')$. Autrement il faut trouver les caractéristiques de la fonction aléatoire de sortie à partir de celles du signal appliqué à l'entrée du système.

Nous allons tout d'abord montrer qu'on peut se limiter au cas d'un opérateur homogène. En effet supposons que l'opérateur L ne soit pas homogène et s'exprime par la formule:

$$L\{X(t)\} = L_0\{X(t)\} + \varphi(t), \qquad (15.7.2)$$

où L_0 est un opérateur linéaire homogène et $\varphi(t)$ une fonction quelconque non aléatoire. On a alors:

$$m_y(t) = M[L_0\{X(t)\}] + \varphi(t), \qquad (15.7.3)$$

c'est-à-dire que la fonction $\varphi(t)$ s'ajoute simplement à l'espérance mathématique de la fonction aléatoire à la sortie du système linéaire. Pour ce qui est de la fonction de corrélation, elle ne change pas lorsqu'on ajoute à la fonction aléatoire un terme non aléatoire.

C'est pourquoi dans l'exposé ultérieur on entendra par opérateurs linéaires des *opérateurs linéaires homogènes*.

Nous allons tout d'abord déterminer les caractéristiques de sortie d'un système linéaire pour certains cas particuliers d'opérateurs.

1. Intégrale d'une fonction aléatoire

Soit une fonction aléatoire $X(t)$ d'espérance mathématique $m_x(t)$ et de fonction de corrélation $K_x(t, t')$. La fonction aléatoire $Y(t)$ est liée à $X(t)$ par l'opérateur homogène d'intégration:

$$Y(t) = \int_0^t X(\tau)\,d\tau. \tag{15.7.4}$$

Trouver les caractéristiques $Y(t)$, à savoir $m_y(t)$ et $K_y(t, t')$.

Ecrivons l'intégrale (15.7.4) comme la limite d'une somme:

$$Y(t) = \int_0^t X(\tau)\,d\tau = \lim_{\Delta\tau \to 0} \sum_i X(\tau_i)\,\Delta\tau, \tag{15.7.5}$$

et trouvons l'espérance mathématique de l'égalité (15.7.5). En vertu du théorème d'addition des espérances mathématiques on a:

$$m_y(t) = M[Y(t)] = \lim_{\Delta\tau \to 0} \sum_i M[X(\tau_i)]\,\Delta\tau =$$

$$= \lim_{\Delta\tau \to 0} \sum_i m_x(\tau_i)\,\Delta\tau = \int_0^t m_x(\tau)\,d\tau\,^{*}), \tag{15.7.6}$$

d'où

$$m_y(t) = \int_0^t m_x(\tau)\,d\tau, \tag{15.7.7}$$

c'est-à-dire que l'*espérance mathématique de l'intégrale d'une fonction aléatoire est égale à l'intégrale de son espérance mathématique*. En d'autres termes on peut intervertir les opérations d'intégration et de recherche de l'espérance mathématique. Ce résultat est tout naturel, car une intégration ne diffère pas dans sa nature d'une sommation qui, comme nous l'avons déjà vu, peut être interchangée avec l'opération d'espérance mathématique.

---

*) On suppose que l'espérance mathématique de la limite est égale à la limite de l'espérance mathématique. En pratique on a généralement affaire à des fonctions pour lesquelles ceci est vrai.

Calculons maintenant la fonction de corrélation $K_y(t, t')$. A cet effet nous passons aux fonctions aléatoires centrées:

$$\overset{\circ}{X}(t) = X(t) - m_x(t); \quad \overset{\circ}{Y}(t) = Y(t) - m_y(t).$$

Il est facile de voir que

$$\overset{\circ}{Y}(t) = \int_0^t \overset{\circ}{X}(\tau)\, d\tau. \tag{15.7.8}$$

Par définition de la fonction de corrélation:

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\, \overset{\circ}{Y}(t')],$$

où

$$\overset{\circ}{Y}(t) = \int_0^t \overset{\circ}{X}(\tau)\, d\tau; \quad \overset{\circ}{Y}(t') = \int_0^{t'} \overset{\circ}{X}(\tau')\, d\tau'\ ^{*)}. \tag{15.7.9}$$

Faisons le produit des expressions (15.7.9):

$$\overset{\circ}{Y}(t)\, \overset{\circ}{Y}(t') = \int_0^t \overset{\circ}{X}(\tau)\, d\tau \int_0^{t'} \overset{\circ}{X}(\tau')\, d\tau'. \tag{15.7.10}$$

Il est facile de voir que le produit des deux intégrales de (15.7.10) est égal à l'intégrale double:

$$\int_0^t \int_0^{t'} \overset{\circ}{X}(\tau)\, \overset{\circ}{X}(\tau')\, d\tau\, d\tau'. \tag{15.7.11}$$

Comme la fonction sous l'intégrale dans (15.7.11) est le produit de deux facteurs dont le premier ne dépend que de τ et le second que de τ', l'intégrale double (15.7.11) se décompose en deux intégrales simples (15.7.10). Par conséquent:

$$\overset{\circ}{Y}(t)\, \overset{\circ}{Y}(t') = \int_0^t \int_0^{t'} \overset{\circ}{X}(\tau)\, \overset{\circ}{X}(\tau')\, d\tau\, d\tau'. \tag{15.7.12}$$

En prenant l'espérance mathématique de l'équation (15.7.12) et en intervertissant cette opération avec celle d'intégration on obtient:

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\, \overset{\circ}{Y}(t')] = \int_0^t \int_0^{t'} M[\overset{\circ}{X}(\tau)\, \overset{\circ}{X}(\tau')]\, d\tau\, d\tau'$$

---

*) On sait que dans une intégrale définie la variable d'intégration peut être désignée par une lettre quelconque; ici il est commode d'utiliser les désignations τ et τ' respectivement.

et finalement

$$K_y(t, t') = \int_0^t \int_0^{t'} K_x(\tau, \tau')\, d\tau\, d\tau'. \qquad (15.7.13)$$

Ainsi pour trouver la fonction de corrélation de l'intégrale d'une fonction aléatoire, il faut *intégrer deux fois* la fonction de corrélation de la fonction aléatoire initiale, d'abord sur l'un des arguments, puis sur l'autre

2. *Dérivée d'une fonction aléatoire*

Soit une fonction aléatoire $X(t)$ d'espérance mathématique $m_x(t)$ et de fonction de corrélation $K_x(t, t')$. La fonction aléatoire $Y(t)$ est liée à la fonction aléatoire $X(t)$ par l'opérateur homogène linéaire de dérivation:

$$Y(t) = \frac{dX(t)}{dt}. \qquad (15.7.14)$$

Il y a lieu de trouver $m_y(t)$ et $K_y(t, t')$.

Ecrivons la dérivée comme la limite:

$$Y(t) = \lim_{\Delta t \to 0} \frac{X(t+\Delta t) - X(t)}{\Delta t}. \qquad (15.7.15)$$

En prenant l'espérance mathématique des deux membres de l'égalité (15.7.15) on obtient:

$$m_y(t) = M[Y(t)] = \lim_{\Delta t \to 0} \frac{m_x(t+\Delta t) - m_x(t)}{\Delta t} = \frac{dm_x(t)}{dt}\ ^{*)},$$

d'où

$$m_y(t) = \frac{dm_x(t)}{dt}, \qquad (15.7.16)$$

c'est-à-dire que l'*espérance mathématique de la dérivée d'une fonction aléatoire est égale à la dérivée de son espérance mathématique.* Par conséquent, l'opération de dérivation tout comme celle d'intégration peut être intervertie avec celle de l'espérance mathématique.

Pour trouver $K_y(t, t')$ nous allons passer aux fonctions aléatoires centrées $\overset{\circ}{Y}(t)$ et $\overset{\circ}{X}(t)$; de toute évidence on a:

$$\overset{\circ}{Y}(t) = \frac{d\overset{\circ}{X}(t)}{dt}. \qquad (15.7.17)$$

Par définition:

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\, \overset{\circ}{Y}(t')].$$

*) Voir note à la page 378.

En remplaçant $\dot{Y}(t)$ et $\dot{Y}(t')$ par leurs expressions, on a :

$$K_y(t, t') = M\left[\frac{d\dot{X}(t)}{dt}\,\frac{d\dot{X}(t')}{dt'}\right].$$

Ecrivons l'expression sous le signe de l'espérance mathématique comme une dérivée partielle mixte, soit :

$$\frac{d\dot{X}(t)}{dt}\,\frac{d\dot{X}(t')}{dt'} = \frac{\partial^2 \dot{X}(t)\,\dot{X}(t')}{\partial t\,\partial t'}. \qquad (15.7.18)$$

Nous avons montré que l'espérance mathématique de la dérivée d'une fonction aléatoire est égale à la dérivée de son espérance mathématique, c'est-à-dire que les signes de dérivation et d'espérance mathématique peuvent être intervertis. Par conséquent :

$$K_y(t, t') = M\left[\frac{\partial^2 \dot{X}(t)\,\dot{X}(t')}{\partial t\,\partial t'}\right] = \frac{\partial^2}{\partial t\,\partial t'} M[\dot{X}(t)\,\dot{X}(t')] = \frac{\partial^2}{\partial t\,\partial t'} K_x(t, t'). \qquad (15.7.19)$$

Par conséquent :

$$K_y(t, t') = \frac{\partial^2 K_x(t, t')}{\partial t\,\partial t'}. \qquad (15.7.20)$$

Ainsi pour trouver la fonction de corrélation d'une dérivée il y a lieu de *dériver deux fois* la fonction initiale successivement par rapport à l'un et à l'autre de ses arguments.

En comparant les règles de recherche de l'espérance mathématique et de la fonction de corrélation pour les deux opérateurs homogènes linéaires envisagés, on voit que ces règles sont analogues. Pour trouver l'espérance mathématique de la fonction aléatoire transformée il y a lieu d'appliquer l'opérateur à l'espérance mathématique de la fonction aléatoire initiale; pour trouver la fonction de corrélation il y a lieu d'appliquer ce même opérateur *deux fois* à la fonction de corrélation de la fonction aléatoire initiale. Dans le premier cas c'était une intégration double, dans le second une double dérivation.

On peut montrer que cette règle est générale pour tous les opérateurs linéaires homogènes *). Ici nous donnons cette règle sans démonstration.

Si un opérateur L transforme une fonction aléatoire $X(t)$ d'espérance mathématique $m_x(t)$ et de fonction de corrélation $K_x(t, t')$ en une autre fonction aléatoire

$$Y(t) = L\{X(t)\},$$

*) Voir, par exemple, V. Pougatchev. *Théorie des fonctions aléatoires et ses applications aux problèmes de la commande automatique* (en russe). Fizmathguiz, Moscou, 1962.

pour trouver l'espérance mathématique de $Y(t)$ il y a lieu d'appliquer le même opérateur à l'espérance mathématique de $X(t)$:

$$m_y(t) = L\{m_x(t)\} \qquad (15.7.21)$$

et pour trouver la fonction de corrélation il faut appliquer deux fois le même opérateur à la fonction de corrélation de $X(t)$, d'abord suivant le premier argument, puis suivant le second :

$$K_y(t, t') = L^{(t)}L^{(t')}\{K_x(t, t')\}. \qquad (15.7.22)$$

Dans la formule (15.7.22) les indices (t), (t') dont est marqué l'opérateur indiquent l'argument suivant lequel il y a lieu de l'appliquer.

Dans de nombreux problèmes, ce qui intéresse en fin de compte ce n'est pas la fonction de corrélation $K_y(t, t')$ à la sortie du système linéaire, mais la variance $D_y(t)$ caractérisant la précision de fonctionnement du système en présence de perturbations aléatoires. On peut trouver la variance $D_y(t)$ à partir de la fonction de corrélation :

$$D_y(t) = K_y(t, t'). \qquad (15.7.23)$$

Il faut ici souligner qu'en général, pour trouver la variance à la sortie d'un système linéaire, il ne suffit pas de connaître la variance à l'entrée, mais aussi la fonction de corrélation. En effet, un système linéaire peut réagir tout à fait différemment à des perturbations aléatoires agissant à son entrée, suivant la structure interne de celles-ci ; suivant, par exemple, qu'elles se composent essentiellement d'oscillations haute ou basse fréquence. Or, la structure interne d'un processus aléatoire est décrite non pas par sa variance, mais par sa fonction de corrélation.

E x e m p l e. L'entrée d'un mécanisme dérivateur est attaquée par une fonction aléatoire $X(t)$ d'espérance mathématique $m_x(t) = \sin t$ et de fonction de corrélation

$$K_x(t, t') = D_x e^{-\alpha(t'-t)^2},$$

où D_x est la variance constante de $X(t)$.

Trouver l'espérance mathématique et la variance à la sortie du système.

S o l u t i o n. La fonction aléatoire $Y(t)$ à la sortie du système (réponse) est liée à l'action $X(t)$ par l'opérateur de dérivation :

$$Y(t) = \frac{d}{dt} X(t).$$

En appliquant les règles générales on obtient :

$$m_y(t) = \frac{d}{dt} m_x(t) = \cos t;$$

$$K_y(t, t') = \frac{\partial^2}{\partial t\, \partial t'} K_x(t, t') = 2D_x\alpha e^{-\alpha(t'-t)^2}[1 - 2\alpha(t'-t)^2].$$

Posant $t' = t$, on a

$$D_y(t) = 2D_x\alpha,$$

ou, en remarquant que $D_y(t)$ ne dépend pas de t,

$$D_y = 2D_x\alpha.$$

Ainsi, la variance à la sortie d'un mécanisme dérivateur dépend non seulement de la variance D_x à l'entrée mais également d'un coefficient α caractérisant la rapidité de décroissance de la corrélation entre les sections de la fonction aléatoire $X(t)$ lorsque l'intervalle les séparant augmente. Si le coefficient α est petit, la corrélation s'amortit lentement, la fonction aléatoire varie dans le temps d'une manière relativement régulière, dans ce cas la dérivation conduit à des erreurs relativement peu importantes. Au contraire si le coefficient α est grand, la fonction de corrélation décroît rapidement ; la fonction aléatoire se compose essentiellement d'oscillations désordonnées à haute fréquence ; la dérivation de cette fonction conduit évidemment à des erreurs aléatoires importantes. Généralement on utilise dans ces cas la méthode de *lissage* de la fonction, à cette fin on modifie l'opérateur du système de sorte à diminuer les erreurs aléatoires à la sortie.

15.8. Addition des fonctions aléatoires

Dans la pratique il est rare que l'entrée du système dynamique soit attaquée par une seule fonction aléatoire $X(t)$; souvent il y en a deux ou plus, chacune d'elles étant due à l'action d'un facteur perturbateur. Il s'agit alors de prendre la somme des fonctions aléatoires ou, plus précisément, de trouver les caractéristiques d'une somme à partir des caractéristiques des composantes.

Ce problème est élémentaire dans le cas où les composantes sont des fonctions aléatoires indépendantes (plus exactement non corrélées) entre elles. Dans le cas général il faut connaître encore une caractéristique, à savoir la *fonction de corrélation mutuelle.*

On appelle fonction de corrélation mutuelle des deux fonctions aléatoires $X(t)$ et $Y(t)$ une fonction certaine de deux arguments t et t' qui pour tout couple de valeurs t, t' est égale à la covariance des sections correspondantes des fonctions aléatoires $X(t)$ et $Y(t)$:

$$R_{xy}(t, t') = M[\overset{\circ}{X}(t)\,\overset{\circ}{Y}(t')]. \quad (15.8.1)$$

La fonction de corrélation mutuelle, tout comme la fonction de corrélation ordinaire, ne change pas lorsqu'on ajoute aux fonctions aléatoires des composantes non aléatoires quelconques et par conséquent lors du centrage des fonctions aléatoires.

Par définition de la fonction de corrélation mutuelle on a :

$$R_{xy}(t, t') = R_{yx}(t', t). \quad (15.8.2)$$

Souvent au lieu de la fonction $R_{xy}(t, t')$ on utilise la fonction de corrélation mutuelle normée:

$$r_{xy}(t, t') = \frac{R_{xy}(t, t')}{\sigma_x(t)\,\sigma_y(t')}. \qquad (15.8.3)$$

Si la fonction de corrélation mutuelle est égale à zéro pour toutes les valeurs de t, t', les fonctions aléatoires $X(t)$ et $Y(t)$ sont dites non corrélées (non liées).

Généralement dans la pratique on établit que des fonctions aléatoires ne sont pas corrélées non pas en se basant sur le fait que leur fonction de corrélation mutuelle est nulle; bien au contraire, on pose cette dernière nulle en partant des considérations physiques suivant lesquelles les fonctions aléatoires étudiées doivent être indépendantes.

Soient, par exemple, deux fonctions aléatoires $X(t)$ et $Y(t)$ correspondant aux oscillations de la tension du secteur en deux points différents, alimentés par des sources différentes, physiquement indépendantes; les deux fonctions aléatoires $X(t)$ et $Y(t)$ sont alors non corrélées. Si les deux points sont alimentés par la même source, il est naturel de supposer que les fonctions $X(t)$ et $Y(t)$ sont corrélées.

En général, si des considérations physiques suggèrent qu'il doit exister une relation entre les fonctions aléatoires figurant dans le problème, il y a lieu d'étudier leurs fonctions de corrélation mutuelles.

Connaissant les espérances mathématiques, les fonctions de corrélation ainsi que la fonction de corrélation mutuelle des fonctions aléatoires $X(t)$ et $Y(t)$, on peut trouver les caractéristiques de leur somme:

$$Z(t) = X(t) + Y(t). \qquad (15.8.4)$$

En vertu du théorème d'addition des espérances mathématiques on a:

$$m_z(t) = m_x(t) + m_y(t), \qquad (15.8.5)$$

donc l'*espérance mathématique de la somme de deux fonctions aléatoires est égale à la somme de leurs espérances mathématiques.*

Pour trouver la fonction de corrélation $K_z(t, t')$ passons aux fonctions aléatoires centrées $\mathring{Z}(t)$, $\mathring{X}(t)$, $\mathring{Y}(t)$. On a de toute évidence:

$$\mathring{Z}(t) = \mathring{X}(t) + \mathring{Y}(t). \qquad (15.8.6)$$

Par définition de la fonction de corrélation:

$$K_z(t, t') = M[\mathring{Z}(t)\,\mathring{Z}(t')] =$$
$$= M[(\mathring{X}(t) + \mathring{Y}(t))(\mathring{X}(t') + \mathring{Y}(t'))] =$$
$$= M[\mathring{X}(t)\,\mathring{X}(t')] + M[\mathring{Y}(t)\,\mathring{Y}(t')] + M[\mathring{X}(t)\,\mathring{Y}(t')] +$$
$$+ M[\mathring{X}(t')\,\mathring{Y}(t)]$$

ou

$$K_z(t, t') = K_x(t, t') + K_y(t, t') + R_{xy}(t, t') + R_{xy}(t', t). \quad (15.8.7)$$

La formule (15.8.7) est analogue à (10.2.7) pour la variance d'une somme de variables aléatoires.

Lorsque les fonctions aléatoires $X(t)$ et $Y(t)$ ne sont pas corrélées, donc $R_{xy}(t, t') \equiv 0$, la formule (15.8.7) devient:

$$K_z(t, t') = K_x(t, t') + K_y(t, t'). \quad (15.8.8)$$

Ainsi la *fonction de corrélation de la somme de deux fonctions aléatoires non corrélées est égale à la somme de leurs fonctions de corrélation.*

Les formules ci-dessus peuvent être généralisées au cas d'un nombre aléatoire de composantes. Si la fonction aléatoire $X(t)$ est la somme de n fonctions aléatoires:

$$X(t) = \sum_{i=1}^{n} X_i(t), \quad (15.8.9)$$

son espérance mathématique est donnée par la formule:

$$m_x(t) = \sum_{i=1}^{n} m_{x_i}(t) \quad (15.8.10)$$

et sa fonction de corrélation par la formule:

$$K_x(t, t') = \sum_{i=1}^{n} K_{x_i}(t, t') + \sum_{i \neq j} R_{x_i x_j}(t, t'), \quad (15.8.11)$$

où la sommation s'étend à toutes les combinaisons possibles des indices i et j deux à deux.

Dans le cas où les fonctions aléatoires $X_i(t)$ sont toutes non corrélées deux à deux, la formule (15.8.11) devient le *théorème d'addition des fonctions de corrélation*:

$$K_x(t, t') = \sum_{i=1}^{n} K_{x_i}(t, t') \quad (15.8.12)$$

qui se lit comme suit: *la fonction de corrélation d'une somme de fonctions aléatoires non corrélées deux à deux est égale à la somme des fonctions de corrélation des composantes.*

La formule (15.8.12) est analogue au théorème d'addition des variances des variables aléatoires.

L'addition d'une fonction aléatoire et d'une variable aléatoire est un cas particulier de l'addition des fonctions aléatoires.

Soient une fonction aléatoire $X(t)$ de caractéristiques $m_x(t)$ et $K_x(t, t')$ et une variable aléatoire Y d'espérance mathématique m_y et de variance D_y. Supposons que $X(t)$ et Y ne soient pas corré-

lées, c'est-à-dire que pour t quelconque on ait :

$$M[\overset{\circ}{X}(t)\overset{\circ}{Y}] = 0.$$

Ajoutons à la fonction aléatoire $X(t)$ la variable aléatoire Y ; le résultat de l'addition est une fonction aléatoire :

$$Z(t) = X(t) + Y. \tag{15.8.13}$$

Calculons les caractéristiques de cette dernière. Il est évident que :

$$m_z(t) = m_x(t) + m_y. \tag{15.8.14}$$

Pour trouver $K_z(t, t')$ considérons la variable aléatoire Y comme un cas particulier de la fonction aléatoire ne variant pas dans le temps et cherchons sa fonction de corrélation :

$$K_y(t, t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')] = M[\overset{\circ}{Y}{}^2] = D_y. \tag{15.8.15}$$

A partir de la formule (15.8.8) d'addition des fonctions de corrélation on obtient :

$$K_z(t, t') = K_x(t, t') + D_y,$$

c'est-à-dire que la *fonction de corrélation de la somme d'une fonction aléatoire et d'une variable aléatoire non corrélée avec cette fonction est égale à la somme de la fonction de corrélation de la fonction aléatoire et d'un terme constant égal à la variance de la variable aléatoire.*

15.9. Fonctions aléatoires complexes

Dans la pratique lorsqu'on utilise l'appareil mathématique des fonctions aléatoires, il est souvent commode d'écrire tant ces dernières que leurs caractéristiques non pas sous forme réelle, mais sous forme complexe. A cet effet il y a lieu de donner une définition d'une *variable aléatoire complexe et d'une fonction aléatoire complexe.*

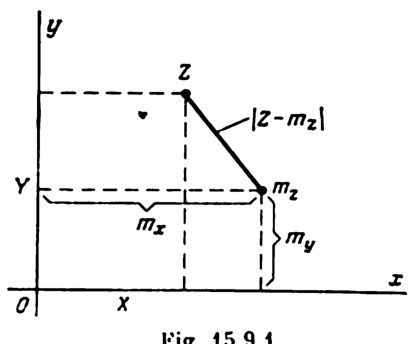


Fig. 15.9.1

On appelle variable aléatoire complexe une variable aléatoire de la forme :

$$Z = X + iY, \tag{15.9.1}$$

où X, Y sont des variables aléatoires réelles ; $i = \sqrt{-1}$ est l'unité imaginaire.

Géométriquement une variable aléatoire complexe peut être interprétée comme un point aléatoire variable Z dans le plan xOy (fig. 15.9.1).

Pour appliquer l'appareil des caractéristiques numériques aux variables aléatoires complexes, il y a lieu de généraliser à ces dernières les notions fondamentales d'espérance mathématique, de variance et de covariance. Il est évident que ces généralisations doivent être telles que dans le cas particulier, où $Y = 0$ et Z réel, elles se réduisent aux caractéristiques habituelles des variables aléatoires.

On appelle *espérance mathématique de la variable aléatoire complexe* $Z = X + iY$ *le nombre complexe*

$$m_z = m_x + im_y. \tag{15.9.2}$$

C'est une certaine valeur moyenne de Z ou, géométriquement, le point moyen m_z autour duquel il y a dispersion du point aléatoire Z (fig. 15.9.1).

On appelle *variance d'une variable aléatoire complexe l'espérance mathématique du carré du module de la variable aléatoire centrée associée*:

$$D_z = M[|\overset{\circ}{Z}|^2], \tag{15.9.3}$$

où

$$\overset{\circ}{Z} = Z - m_z.$$

Du point de vue géométrique la variance d'une variable aléatoire complexe n'est rien d'autre que la valeur moyenne du carré de la distance du point aléatoire Z à son espérance mathématique m_z (fig. 15.9.1). Cette grandeur caractérise la dispersion du point aléatoire Z autour de sa position moyenne.

Exprimons la variance d'une variable aléatoire complexe en fonction de ces parties réelle et imaginaire. De toute évidence on a:

$$\overset{\circ}{Z} = Z - m_z = X + iY - m_x - im_y = \overset{\circ}{X} + i\overset{\circ}{Y},$$

d'où

$$D_z = M[|\overset{\circ}{Z}|^2] = M[\overset{\circ}{X}{}^2 + \overset{\circ}{Y}{}^2] = M[\overset{\circ}{X}{}^2] + M[\overset{\circ}{Y}{}^2]$$

ou

$$D_z = D_x + D_y, \tag{15.9.4}$$

c'est-à-dire que la *variance d'une variable aléatoire complexe est égale à la somme des variances de ses parties réelle et imaginaire.*

Par définition, la variance d'une variable aléatoire complexe est toujours réelle et essentiellement positive. Elle ne peut s'annuler que si la variable Z n'est pas aléatoire.

Evidemment les définitions ci-dessus de l'espérance mathématique et de la variance satisfont à la condition imposée de devenir pour $Y = 0$ et $Z = X$ les définitions habituelles de l'espérance mathématique et de la variance d'une variable aléatoire réelle.

Essayons maintenant de donner une définition analogue aux covariances de deux variables aléatoires complexes Z_1 et Z_2

$$Z_1 = X_1 + iY_1;\ Z_2 = X_2 + iY_2. \tag{15.9.5}$$

Cette définition doit évidemment être telle que pour $Z_1 = Z_2 = Z$ la covariance devienne la variance de Z. On ne peut pas satisfaire à cette condition en définissant, comme dans le cas de variables réelles, la covariance comme l'espérance mathématique du produit $\mathring{Z}_1\mathring{Z}_2$. Il est facile de voir que pour $Z_1 = Z_2 = Z$ l'espérance mathématique de ce produit sera non pas réelle mais complexe, cependant la variance est réelle et essentiellement positive par définition. Pour surmonter cette difficulté on définit la covariance comme l'espérance mathématique du produit de $\mathring{Z}_1$ par le *complexe conjugué* de $\mathring{Z}_2$:

$$\bar{\mathring{Z}}_2 = \mathring{X}_2 - i\mathring{Y}_2. \qquad (15.9.6)$$

Pour $Z_1 = Z_2 = Z$ la covariance deviendra évidemment la variance de Z:

$$K_{zz} = M[(\mathring{X} + i\mathring{Y})(\mathring{X} - i\mathring{Y})] = M[\mathring{X}^2] + M[\mathring{Y}^2] = D_z. \qquad (15.9.7)$$

Il paraît donc raisonnable de s'arrêter sur la définition suivante de la covariance de deux variables aléatoires complexes Z_1 et Z_2:

$$K_{z_1z_2} = M[\mathring{Z}_1\bar{\mathring{Z}}_2], \qquad (15.9.8)$$

où le trait au-dessus de la lettre indique le complexe conjugué de la variable aléatoire respective.

Exprimons la covariance de deux variables aléatoires complexes en fonction des covariances de leurs parties réelles et imaginaires. On a ainsi:

$$K_{z_1z_2} = M[\mathring{Z}_1\bar{\mathring{Z}}_2] = M[(\mathring{X}_1 + i\mathring{Y}_1)(\mathring{X}_2 - i\mathring{Y}_2)] =$$
$$= K_{x_1x_2} + K_{y_1y_2} + i(K_{y_1x_2} - K_{x_1y_2}), \qquad (15.9.9)$$

où $K_{x_1x_2}$, $K_{y_1y_2}$, $K_{y_1x_2}$, $K_{x_1y_2}$ sont les covariances des couples de variables (X_1, X_2), (Y_1, Y_2); (Y_1, X_2), (X_1, Y_2) respectivement.

Il est évident que lorsque toutes ces variables ne sont pas corrélées entre elles, la covariance des variables Z_1, Z_2 est également nulle.

Résumons les particularités qui différencient les définitions habituelles des caractéristiques des variables aléatoires réelles de celles des variables aléatoires complexes:

1) la variance n'est plus l'espérance mathématique du carré de la variable aléatoire centrée, mais l'espérance mathématique du *carré de son module*;

2) la covariance ne se définit plus comme l'espérance mathématique du produit des variables centrées, mais comme celle du produit d'une variable centrée par le *complexe conjugué* de l'autre.

Nous allons passer maintenant à la définition d'une fonction aléatoire complexe et de ses caractéristiques.

On appelle fonction aléatoire complexe une fonction de la forme:

$$Z(t) = X(t) + iY(t), \tag{15.9.10}$$

où $X(t)$, $Y(t)$ sont des fonctions aléatoires réelles.

L'espérance mathématique de la fonction aléatoire complexe (15.9.10) est égale à:

$$m_z(t) = m_x(t) + im_y(t). \tag{15.9.11}$$

La variance de la fonction aléatoire complexe $Z(t)$ est l'espérance mathématique du carré du module de la fonction centrée correspondante:

$$D_z(t) = M[|\mathring{Z}(t)|^2], \tag{15.9.12}$$

où

$$\mathring{Z}(t) = Z(t) - m_z(t) = \mathring{X}(t) + i\mathring{Y}(t). \tag{15.9.13}$$

On déduit de (15.9.12) que la variance d'une fonction aléatoire complexe est réelle et non négative.

En vertu de la formule (15.9.4) la variance d'une fonction aléatoire complexe est égale à la somme des variances de ses parties réelle et imaginaire:

$$D_z(t) = D_x(t) + D_y(t). \tag{15.9.14}$$

La fonction de corrélation d'une fonction aléatoire complexe est la covariance de ses sections t et t':

$$K_z(t, t') = M[\mathring{Z}(t)\overline{\mathring{Z}}(t')], \tag{15.9.15}$$

où

$$\overline{\mathring{Z}}(t') = \mathring{X}(t') - i\mathring{Y}(t')$$

est la fonction complexe conjuguée de $\mathring{Z}(t')$.

Pour $t' = t$ la fonction de corrélation devient évidemment une variance:

$$K_z(t, t) = D_z(t). \tag{15.9.16}$$

En utilisant la formule (15.9.9) on peut exprimer la fonction de corrélation d'une fonction aléatoire complexe à l'aide des caractéristiques de ses parties réelle et imaginaire. Si les fonctions Z_1 et Z_2 figurant dans cette formule sont les sections $Z(t)$ et $Z(t')$ d'une fonction aléatoire, on obtient:

$$K_z(t, t') = K_x(t, t') + K_y(t, t') + i\{R_{xy}(t', t) - R_{xy}(t, t')\}, \tag{15.9.17}$$

où $R_{xy}(t, t')$ est la fonction de corrélation mutuelle des fonctions aléatoires $X(t)$ et $Y(t)$ (parties réelle et imaginaire de la fonction $Z(t)$).

Dans le cas où les parties réelle et imaginaire d'une fonction aléatoire ne sont pas corrélées $[R_{xy}(t, t') \equiv 0]$, la formule (15.9.17) devient :

$$K_z(t, t') = K_x(t, t') + K_y(t, t'). \tag{15.9.18}$$

Dans l'exposé ultérieur nous utiliserons tant la forme réelle que la forme complexe des fonctions aléatoires. Dans ce dernier cas nous en ferons toujours mention.

CHAPITRE 16

DÉVELOPPEMENTS CANONIQUES DES FONCTIONS ALÉATOIRES

16.1. Méthode des développements canoniques. Représentation de la fonction aléatoire sous la forme d'une somme de fonctions aléatoires élémentaires

Dans le § 15.7 nous avons étudié les règles générales des transformations linéaires des fonctions aléatoires. Ainsi, lors d'une transformation linéaire d'une fonction aléatoire, son espérance mathématique subit la même transformation linéaire, et sa fonction de corrélation se trouve soumise à cette transformation deux fois, par rapport à l'un et l'autre de ses arguments.

La règle de transformation de l'espérance mathématique est assez simple et ne présente aucune difficulté lors de son utilisation. Pour ce qui est de la double transformation de la fonction de corrélation, elle conduit souvent à des calculs fort compliqués, ce qui rend difficile l'emploi pratique des méthodes générales examinées.

En effet, soit l'exemple simple de l'opérateur intégral:

$$Y(t)=\int_0^t X(\tau)\,d\tau. \qquad (16.1.1)$$

En vertu de la règle générale la fonction de corrélation est transformée deux fois par le même opérateur:

$$K_y(t,t')=\int_0^t\int_0^{t'} K_x(\tau,\tau')\,d\tau\,d\tau'. \qquad (16.1.2)$$

Très souvent la fonction de corrélation $K_x(t, t')$ obtenue expérimentalement ne peut s'exprimer sous forme analytique et se trouve donnée par un tableau; dans ce cas l'intégrale (16.1.2) doit être calculée numériquement comme une fonction de deux limites. Ce procédé est très long et compliqué. Même si on arrive à trouver une approximation analytique pour la fonction sous l'intégrale, dans la majorité des cas l'intégrale (16.1.2) ne peut s'écrire à l'aide de fonctions connues. Il en est ainsi même pour les opérateurs de transformation les plus simples. Si, comme c'est souvent le cas, le fonctionnement d'un système dynamique est décrit par des équations différentielles dont la solution ne peut s'exprimer sous forme explicite, la déter-

mination de la fonction de corrélation à la sortie est encore plus compliquée, car dans ce cas il y a lieu d'intégrer des équations différentielles aux dérivées partielles.

C'est pourquoi, dans la pratique, les méthodes générales de transformation des fonctions aléatoires se trouvent être trop compliquées pour qu'il soit raisonnable de les employer. Dans les applications pratiques le plus souvent on utilise des méthodes conduisant à des

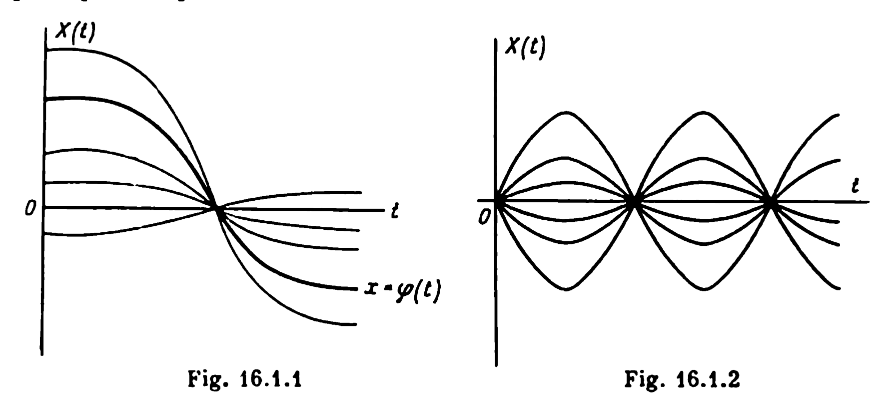


Fig. 16.1.1

Fig. 16.1.2

transformations plus simples. L'une de ces méthodes, proposée par V. Pougatchev et portant le nom de *méthode des développements canoniques*, fait l'objet du présent chapitre.

La méthode des développements canoniques repose essentiellement sur la représentation de la fonction aléatoire à transformer sous la forme d'une somme de *fonctions aléatoires élémentaires*.

On appelle fonction aléatoire élémentaire une fonction de la forme:

$$X(t) = V\varphi(t), \tag{16.1.3}$$

où V est une variable aléatoire, $\varphi(t)$ une fonction non aléatoire.

La fonction aléatoire élémentaire est la forme la plus simple des fonctions aléatoires. En effet, dans l'expression (16.1.3) seul le facteur V est aléatoire, tandis que la fonction du temps $\varphi(t)$ ne l'est pas.

Toutes les réalisations possibles de la fonction aléatoire élémentaire $X(t)$ peuvent être obtenues à partir de la courbe $x = \varphi(t)$ par changement d'échelle suivant l'axe des ordonnées (fig. 16.1.1). L'axe des abscisses ($x = 0$) se trouve être l'une des réalisations possibles de la fonction aléatoire $X(t)$, ce cas ayant lieu lorsque $V = 0$ (si seulement zéro figure parmi les valeurs possibles de V).

A titre d'exemples de fonctions aléatoires élémentaires citons $X(t) = V \sin t$ (fig. 16.1.2) et $X(t) = Vt^2$ (fig. 16.1.3).

Les fonctions aléatoires élémentaires ont ceci de particulier que les deux propriétés essentielles des fonctions aléatoires s'y trouvent

séparées, le hasard n'intervenant que dans le coefficient V et la variation dans le temps étant exprimée par une fonction certaine $\varphi(t)$.

Déterminons les caractéristiques de la fonction aléatoire élémentaire (16.1.3). On a:

$$m_x(t) = M[V\varphi(t)] = m_v\varphi(t),$$

où m_v est l'espérance mathématique de la variable aléatoire V.

Si $m_v = 0$, l'espérance mathématique de la fonction aléatoire $X(t)$ est également nulle; de plus on a:

$$m_x(t) \equiv 0.$$

Toute fonction aléatoire peut être centrée, c'est-à-dire écrite sous une forme telle que son espérance mathématique soit nulle. C'est pourquoi ultérieurement *nous n'envisagerons que des fonctions aléatoires élémentaires centrées* pour lesquelles $m_v = 0$; $V = \overset{\circ}{V}$; $m_x(t) \equiv 0$.

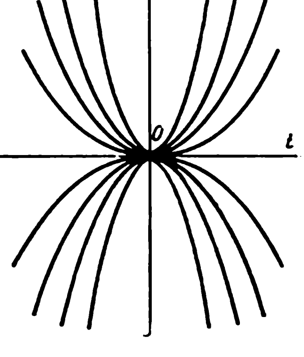


Fig. 16.1.3

Déterminons la fonction de corrélation d'une fonction aléatoire élémentaire $X(t)$. On a:

$$K_x(t, t') = M[X(t)X(t')] = \\ = \varphi(t)\varphi(t')M[V^2] = \varphi(t)\varphi(t')D,$$

où D est la variance de V.

Des fonctions aléatoires élémentaires se prêtent facilement à diverses transformations linéaires.

Dérivons, par exemple, la fonction aléatoire (16.1.3). La variable aléatoire V ne dépendant pas de t doit être mise en facteur devant le signe de dérivation; on a alors:

$$X'(t) = V\varphi'(t).$$

D'une manière analogue on a:

$$\int_0^t X(\tau)\,d\tau = V\int_0^t \varphi(\tau)\,d\tau.$$

En général, si la fonction aléatoire élémentaire est transformée par un opérateur linéaire L, seule la fonction non aléatoire $\varphi(t)$ se trouve transformée par l'opérateur L, le facteur aléatoire V, comme indépendant du temps, devant être sorti du signe de l'opérateur:

$$L\{X(t)\} = VL\{\varphi(t)\}. \tag{16.1.4}$$

Ceci signifie que si une fonction aléatoire élémentaire est appliquée à l'entrée d'un système linéaire, seule la fonction non aléatoire $\varphi(t)$ subit cette transformation. Ceci nous amène à l'idée de repré-

senter, exactement ou approximativement, la fonction aléatoire du type général attaquant l'entrée d'un système dynamique comme la somme de fonctions aléatoires élémentaires et seulement après la soumettre à la transformation. Cette idée de décomposition d'une fonction aléatoire en une somme de fonctions aléatoires élémentaires est à la base de la méthode des développements canoniques.

Soit la fonction aléatoire:

$$X(t) = m_x(t) + \overset{\circ}{X}(t). \tag{16.1.5}$$

Supposons que d'une manière exacte ou approximative, cette fonction puisse être écrite comme une somme:

$$X(t) = m_x(t) + \sum_{i=1}^{m} V_i \varphi_i(t), \tag{16.1.6}$$

où V_i sont des variables aléatoires d'espérances mathématiques nulles; $\varphi_i(t)$ sont des fonctions non aléatoires; $m_x(t)$ est l'espérance mathématique de la fonction $X(t)$.

La formule (16.1.6) est le *développement de la fonction aléatoire* $X(t)$. Les variables aléatoires $V_1, V_2, \ldots, V_m$ sont les *coefficients* du développement, et les fonctions non aléatoires $\varphi_1(t), \varphi_2(t), \ldots$ $\ldots, \varphi_m(t)$ les *fonctions de coordonnées*.

Déterminons maintenant la réponse d'un système linéaire d'opérateur L à une action aléatoire $X(t)$ donnée par le développement (16.1.6). On sait qu'un système linéaire est doué de la *propriété de superposition* qui consiste en ce que la réponse du système à une somme de plusieurs actions est égale à la somme des réponses à chaque action particulière. En effet, étant linéaire, l'opérateur L du système peut, par définition, être appliqué à tous les termes séparément.

Désignant par $Y(t)$ la réponse du système à l'action aléatoire $X(t)$ on a:

$$Y(t) = L\{X(t)\} = L\{m_x(t)\} + \sum_{i=1}^{m} V_i L\{\varphi_i(t)\}. \tag{16.1.7}$$

Ecrivons l'expression (16.1.7) sous une forme un peu différente. Compte tenu de la règle de transformation linéaire de l'espérance mathématique on a:

$$L\{m_x(t)\} = m_y(t).$$

En introduisant la désignation:

$$L\{\varphi_i(t)\} = \psi_i(t),$$

on obtient:

$$Y(t) = m_y(t) + \sum_{i=1}^{m} V_i \psi_i(t). \tag{16.1.8}$$

L'expression (16.1.8) n'est rien d'autre que le *développement* de la fonction aléatoire $Y(t)$ suivant des fonctions élémentaires. Les

coefficients de ce développement sont les mêmes variables aléatoires $V_1, V_2, \ldots, V_m$ qu'en (16.1.6), tandis que l'espérance mathématique et les fonctions de coordonnées sont obtenues à partir de l'espérance mathématique et des fonctions de coordonnées de la fonction aléatoire initiale moyennant la même transformation L à laquelle a été soumise la fonction aléatoire $X(t)$.

On obtient donc la règle suivante de transformation des fonctions aléatoires données sous la forme d'un développement.

Lorsque la fonction aléatoire $X(t)$ donnée par un développement suivant les fonctions élémentaires est soumise à une transformation linéaire L, les coefficients du développement restent inchangés et l'espérance mathématique et les fonctions de coordonnées subissent la même transformation linéaire L.

Ainsi, l'avantage essentiel que présente la décomposition d'une fonction aléatoire en fonctions élémentaires est en ce qu'une transformation linéaire d'une fonction aléatoire se réduit aux mêmes transformations linéaires des fonctions non aléatoires, à savoir de l'espérance mathématique et des fonctions de coordonnées. Ceci simplifie considérablement le problème de la détermination des caractéristiques de la fonction aléatoire $Y(t)$ par rapport à la solution générale donnée au § 15.7. En effet, chacune des fonctions non aléatoires $m_x(t)$, $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_m(t)$ n'est transformée qu'une seule fois à la différence de la fonction de corrélation $K_x(t, t')$ qui, en vertu des règles générales, subit une double transformation.

16.2. Développement canonique d'une fonction aléatoire

Soit la fonction aléatoire $X(t)$ donnée par le développement:

$$X(t) = m_x(t) + \sum_{i=1}^{m} V_i\varphi_i(t), \qquad (16.2.1)$$

où les coefficients $V_1, V_2, \ldots, V_m$ forment un système de variables aléatoires d'espérances mathématiques nulles et de matrice de corrélation $\| K_{ij} \|$.

Cherchons la fonction de corrélation et la variance de la fonction aléatoire $X(t)$.

Par définition on a:

$$K_x(t, t') = M[\mathring{X}(t)\mathring{X}(t')], \qquad (16.2.2)$$

où

$$\mathring{X}(t) = \sum_{i=1}^{m} V_i\varphi_i(t), \qquad (16.2.3)$$

$$\mathring{X}(t') = \sum_{j=1}^{m} V_j\varphi_j(t'). \qquad (16.2.4)$$

Dans la formule (16.2.4) l'indice de sommation est désigné par j, ceci pour souligner son indépendance de l'indice de sommation i dans la formule (16.2.3).

En multipliant les expressions (16.2.3) et (16.2.4) et en appliquant l'opération d'espérance mathématique au produit on obtient :

$$K_x(t, t') = M\left[\sum_{i, j} V_i V_j \varphi_i(t)\varphi_j(t')\right] = \\ = \sum_{i, j} M[V_i V_j]\varphi_i(t)\varphi_j(t'), \qquad (16.2.5)$$

où la sommation s'étend à tous les couples de valeurs i, j, qu'elles soient égales entre elles ou non. Dans le cas où $i = j$ on a :

$$M[V_i V_j] = M[V_i^2] = K_{ii} = D_i,$$

D_i étant la variance de la variable aléatoire V_i. Quand $i \neq j$, on a :

$$M[V_i V_j] = K_{ij},$$

K_{ij} étant la covariance des variables aléatoires V_i, V_j.

Substituons ces valeurs dans la formule (16.2.5), on obtient l'expression de la fonction de corrélation de $X(t)$ donnée par le développement (16.2.1), il vient :

$$K_x(t, t') = \sum_{i=1}^{m} \varphi_i(t)\varphi_i(t')D_i + \sum_{i\neq j} \varphi_i(t)\varphi_j(t')K_{ij}. \qquad (16.2.6)$$

En posant $t' = t$ dans l'expression (16.2.6) on obtient la variance de la fonction aléatoire $X(t)$:

$$D_x(t) = \sum_{i=1}^{m} [\varphi_i(t)]^2 D_i + \sum_{i\neq j} \varphi_i(t)\varphi_j(t)K_{ij}. \qquad (16.2.7)$$

Il est évident que les expressions (16.2.6) et (16.2.7) deviennent particulièrement simples lorsque tous les coefficients V_i du développement (16.2.1) ne sont pas corrélés, c'est-à-dire $K_{ij} = 0$ pour $i \neq j$. Dans ce cas le développement de la fonction aléatoire est dit *canonique*.

On appelle *développement canonique* de la fonction aléatoire $X(t)$ sa représentation sous la forme :

$$X(t) = m_x(t) + \sum_{i=1}^{m} V_i \varphi_i(t), \qquad (16.2.8)$$

où $m_x(t)$ est l'espérance mathématique de la fonction aléatoire ; $\varphi_1(t)$, $\varphi_2(t)$, . . ., $\varphi_m(t)$ les fonctions de coordonnées, et V_1, V_2, ..., V_m des variables aléatoires non corrélées d'espérances mathématiques nulles.

Si l'on connaît le développement canonique d'une fonction aléatoire, sa fonction de corrélation s'exprime d'une manière fort simple.

En posant $K_{ij} = 0$ pour $i \neq j$ dans la formule (16.2.6) on obtient :

$$K_x(t, t') = \sum_{i=1}^{m} \varphi_i(t)\, \varphi_i(t')\, D_i. \tag{16.2.9}$$

L'expression (16.2.9) est appelée *développement canonique de la fonction de corrélation.*

En posant $t' = t$ dans la formule (16.2.9) on obtient la variance de la fonction aléatoire :

$$D_x(t) = \sum_{i=1}^{m} [\varphi_i(t)]^2 D_i. \tag{16.2.10}$$

Ainsi, connaissant le développement canonique de la fonction aléatoire $X(t)$ on peut trouver directement le développement canonique de sa fonction de corrélation. On peut montrer que l'inverse est également vrai, à savoir si le développement canonique de la fonction de corrélation (16.2.9) est donné, le développement canonique de la fonction aléatoire $X(t)$ sera de la forme (16.2.8) avec les fonctions de coordonnées $\varphi_i(t)$ et les coefficients V_i de variances D_i. Nous laissons cette assertion sans démonstration spéciale *).

Le nombre de termes du développement canonique d'une fonction aléatoire peut être tant fini qu'infini. Dans le chapitre 17 nous rencontrerons des exemples de développements canoniques à un nombre infini de termes. De plus dans certains cas on utilise des *représentations canoniques intégrales* des fonctions aléatoires, où la somme est remplacée par une intégrale.

Les développements canoniques peuvent être appliqués aussi bien aux fonctions aléatoires réelles qu'aux fonctions complexes. Voyons comment on peut généraliser la notion de développement canonique au cas des fonctions aléatoires complexes.

On appelle fonction élémentaire complexe une fonction de la forme :

$$X(t) = V\varphi(t), \tag{16.2.11}$$

où la variable aléatoire V et la fonction $\varphi(t)$ sont complexes.

Déterminons la fonction de corrélation de la fonction aléatoire élémentaire (16.2.11). En utilisant la définition générale de la fonction de corrélation d'une fonction aléatoire complexe on obtient :

$$K_x(t, t') = M[V\varphi(t)\overline{V\varphi(t')}], \tag{16.2.12}$$

où le trait indique comme auparavant une grandeur complexe conjuguée. Compte tenu de

$$\overline{V\varphi(t)} = \overline{V}\,\overline{\varphi(t)}$$

*) Pour la démonstration voir : V. Pougatchev. *Théorie des fonctions aléatoires et ses applications aux problèmes de la commande automatique* (en russe). Fizmathguiz, Moscou, 1962.

et en mettant en facteur devant le signe de l'espérance mathématique les fonctions certaines $\varphi(t)$ et $\overline{\varphi(t')}$ on obtient :

$$K_x(t, t') = \varphi(t)\overline{\varphi(t')}M[|V|^2].$$

Mais en vertu du § 15.9, le terme $M[|V|^2]$ n'est rien d'autre que la variance de la variable aléatoire complexe V, donc :

$$M[|V|^2] = D,$$

par conséquent :

$$K_x(t, t') = \varphi(t)\overline{\varphi(t')}D. \tag{16.2.13}$$

Le développement canonique d'une fonction aléatoire complexe s'écrit sous la forme :

$$X(t) = m_x(t) + \sum_{i=1}^{m} V_i\varphi_i(t), \tag{16.2.14}$$

où $V_1, V_2, \ldots, V_m$ sont des variables aléatoires complexes non corrélées d'espérances mathématiques nulles, et $m_x(t)$, $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_m(t)$ des fonctions non aléatoires complexes.

Si une fonction aléatoire complexe est donnée par le développement canonique (16.2.14), sa fonction de corrélation s'écrit :

$$K_x(t, t') = \sum_{i=1}^{m} \varphi_i(t)\overline{\varphi_i(t')}D_i. \tag{16.2.15}$$

où D_i est la variance de V_i :

$$D_i = M[|V_i|^2]. \tag{16.2.16}$$

La formule (16.2.15) découle directement de l'expression (16.2.13) pour la fonction de corrélation d'une fonction aléatoire élémentaire complexe.

L'expression (16.2.15) est appelée développement canonique de la fonction de corrélation d'une fonction aléatoire complexe.

Posant $t' = t$ dans (16.2.15) on obtient l'expression de la variance d'une fonction aléatoire complexe donnée par le développement (16.2.14) :

$$D_x(t) = \sum_{i=1}^{m} |\varphi_i(t)|^2 D_i. \tag{16.2.17}$$

16.3. Transformations linéaires des fonctions aléatoires données par leurs développements canoniques

Soit la fonction aléatoire $X(t)$ appliquée à l'entrée d'un système linéaire L (fig. 16.3.1).

Le système fait subir à la fonction $X(t)$ la transformation linéaire L de sorte qu'on obtient à la sortie la fonction aléatoire :

$$Y(t) = L\{X(t)\}. \tag{16.3.1}$$

Supposons que la fonction aléatoire $X(t)$ soit donnée par son développement canonique:

$$X(t) = m_x(t) + \sum_{i=1}^{k} V_i \varphi_i(t). \qquad (16.3.2)$$

Déterminons la réponse du système à cette action. L'opérateur du système étant linéaire, on a:

$$Y(t) = L\{X(t)\} = L\{m_x(t)\} + \sum_{i=1}^{k} V_i L\{\varphi_i(t)\}. \qquad (16.3.3)$$

Il est facile de voir que l'expression (16.3.3) n'est rien d'autre que le *développement canonique de la fonction aléatoire* $Y(t)$ dont l'espé-

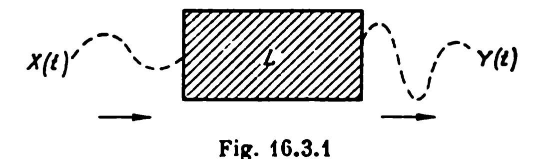


Fig. 16.3.1

rance mathématique est:

$$m_y(t) = L\{m_x(t)\} \qquad (16.3.4)$$

et les fonctions de coordonnées sont:

$$\psi_i(t) = L\{\varphi_i(t)\}. \qquad (16.3.5)$$

Ainsi lors d'une transformation linéaire du développement canonique de la fonction aléatoire $X(t)$ on obtient le développement canonique de la fonction aléatoire $Y(t)$, l'espérance mathématique et les fonctions de coordonnées étant soumises à la même transformation.

Si une fonction aléatoire $Y(t)$ s'obtient à la sortie d'un système linéaire sous la forme d'un développement canonique:

$$Y(t) = m_y(t) + \sum_{i=1}^{k} V_i \psi_i(t), \qquad (16.3.6)$$

sa fonction de corrélation et sa variance se déterminent facilement par les formules:

$$K_y(t, t') = \sum_{i=1}^{k} \psi_i(t)\, \psi_i(t')\, D_i, \qquad (16.3.7)$$

$$D_y(t) = \sum_{i=1}^{k} [\psi_i(t)]^2 D_i. \qquad (16.3.8)$$

C'est là l'avantage essentiel des développements *canoniques* par rapport aux autres modes de développement suivant des fonctions élémentaires.

Examinons plus en détail l'application de la méthode des développements canoniques au calcul de la réponse d'un système dynamique à une action aléatoire d'entrée, dont le fonctionnement est décrit par une équation différentielle linéaire, dans le cas général, à coefficients variables. Ecrivons cette équation sous la forme opérationnelle:

$$A_n(p, t)\, Y(t) = B_m(p, t)\, X(t). \tag{16.3.9}$$

En vertu des règles des transformations linéaires des fonctions aléatoires exposées ci-dessus, les espérances mathématiques de l'action et la réponse doivent satisfaire à la même équation:

$$A_n(p, t)\, m_y(t) = B_m(p, t)\, m_x(t). \tag{16.3.10}$$

D'une manière analogue, chacune des fonctions de coordonnées doit vérifier la même équation différentielle:

$$A_n(p, t)\, \psi_i(t) = B_m(p, t)\, \varphi_i(t), \quad (i = 1, 2, \ldots, k). \tag{16.3.11}$$

Ainsi, le problème de la détermination de la réponse d'un système dynamique linéaire à une action aléatoire se réduit au problème mathématique courant de l'intégration de $k + 1$ équations différentielles ordinaires contenant des fonctions non aléatoires. Comme le problème fondamental de l'analyse d'un système dynamique, qui consiste dans la détermination de la réponse à une action donnée, nécessite l'intégration de l'équation différentielle décrivant le fonctionnement du système, et que cette intégration est faite selon une méthode ou une autre, la résolution des équations (16.3.10) et (16.3.11) n'entraîne pas de nouvelles difficultés mathématiques. En particulier, pour la résolution de ces équations on peut utiliser les mêmes dispositifs intégrateurs ou simulateurs que ceux dont on se sert dans l'analyse du fonctionnement des systèmes sans perturbations aléatoires.

Il ne reste plus qu'à élucider les questions relatives aux conditions initiales pour lesquelles il faut intégrer les équations (16.3.10) et (16.3.11).

Nous allons étudier tout d'abord le cas le plus simple lorsque les conditions initiales pour le système dynamique considéré ne sont pas aléatoires. Dans ce cas pour $t = 0$ on doit avoir les conditions suivantes:

$$\left.\begin{array}{r} Y(0) = y_0, \\ Y'(0) = y_1, \\ \cdots\cdots\cdots \\ Y^{(r)}(0) = y_r, \\ \cdots\cdots\cdots \\ Y^{(n-1)}(0) = y_{n-1}, \end{array}\right\} \tag{16.3.12}$$

où $y_0, y_1, \ldots, y_{n-1}$ sont des nombres non aléatoires.

Les conditions (16.3.12) peuvent être écrites sous une forme condensée comme:

$$Y^{(r)}(0)=y_r \quad (r=0,\ 1,\ \ldots,\ n-1), \tag{16.3.13}$$

la « dérivée d'ordre zéro » $Y^{(0)}(t)$ étant ici la fonction $Y(t)$.

Cherchons les conditions initiales pour lesquelles il y a lieu d'intégrer les équations (16.3.10) et (16.3.11). Ecrivons la dérivée d'ordre r de la fonction $Y(t)$ en posant $t=0$:

$$Y^{(r)}(0)=m_y^{(r)}(0)+\sum_{i=1}^{k} V_i\psi_i^{(r)}(0).$$

Compte tenu de (16.3.12) on a:

$$m_y^{(r)}(0)+\sum_{i=1}^{k} V_i\psi_i^{(r)}(0)=y_r. \tag{16.3.14}$$

La grandeur y_r n'étant pas aléatoire, la variance du premier membre de (16.3.14) doit être nulle:

$$\sum_{i=1}^{k} D_i\,[\psi_i^{(r)}(0)]^2=0. \tag{16.3.15}$$

Toutes les variances D_i des variables V_i étant positives, l'égalité (16.3.15) ne peut avoir lieu que si

$$\psi_i^{(r)}(0)=0 \tag{16.3.16}$$

pour tous les i.

En substituant $\psi_i^{(r)}(0)=0$ dans la formule (16.3.14) on obtient:

$$m_y^{(r)}(0)=y_r. \tag{16.3.17}$$

L'égalité (16.3.17) montre que l'équation (16.3.10) pour l'espérance mathématique doit être intégrée pour les conditions initiales (16.3.12), à savoir:

$$\left.\begin{aligned} m_y(0)&=y_0,\\ m_y'(0)&=y_1,\\ &\ldots\ldots\\ m_y^{(r)}(0)&=y_r,\\ &\ldots\ldots\\ m_y^{(n)}(0)&=y_n. \end{aligned}\right\} \tag{16.3.18}$$

Quant aux équations (16.3.11), elles doivent être intégrées *pour des conditions initiales nulles*:

$$\psi_i(0)=\psi_i'(0)=\ldots=\psi_i^{(r)}(0)=\ldots=\psi_i^{(n)}(0)=0 \quad (i=1,\ 2,\ \ldots,\ k). \tag{16.3.19}$$

Envisageons maintenant le cas plus compliqué où les conditions initiales sont aléatoires,

$$\left.\begin{array}{l} Y(0)=Y_0, \\ Y'(0)=Y_1, \\ \dots\dots\dots \\ Y^{(r)}(0)=Y_r, \\ \dots\dots\dots \\ Y^{n-1}(0)=Y_{n-1}, \end{array}\right\} \qquad (16.3.20)$$

$Y_0, Y_1, \ldots, Y_{n-1}$ étant des variables aléatoires.

Dans ce cas la réponse du système s'obtient sous la forme d'une somme:

$$Y(t) = Y_{\mathrm{I}}(t) + Y_{\mathrm{II}}(t), \qquad (16.3.21)$$

où $Y_{\mathrm{I}}(t)$ est la solution de l'équation différentielle (16.3.9) pour des conditions initiales nulles et $Y_{\mathrm{II}}(t)$ la solution de cette même équation différentielle pour le second membre nul et les conditions initiales (16.3.20). La théorie des équations différentielles nous apprend que cette solution doit être une combinaison linéaire des conditions initiales, à savoir

$$Y_{\mathrm{II}}(t) = \sum_{j=0}^{n-1} Y_j f_j(t), \qquad (16.3.22)$$

$f_j(t)$ étant des fonctions non aléatoires.

La solution $Y_{\mathrm{I}}(t)$ peut être obtenue par la méthode exposée ci-dessus sous la forme d'un développement canonique. La fonction de corrélation de la fonction aléatoire $Y(t) = Y_{\mathrm{I}}(t) + Y_{\mathrm{II}}(t)$ se détermine par les méthodes habituelles d'addition des fonctions aléatoires (voir § 15.8).

Notons que souvent dans la pratique pour des instants suffisamment éloignés de l'origine du processus aléatoire, les conditions initiales n'ont aucune influence sur le cours du processus, car les processus transitoires engendrés par celles-ci ont le temps de s'amortir. Les systèmes doués de cette propriété sont dits asymptotiquement stables. Si l'on s'intéresse à la réponse d'un système dynamique asymptotiquement stable sur les intervalles de temps suffisamment éloignés de l'origine, on peut se limiter à la solution $Y_{\mathrm{I}}(t)$ obtenue pour des conditions initiales nulles. Pour des instants suffisamment éloignés de l'instant initial cette solution sera vraie pour des conditions initiales quelconques.

CHAPITRE 17

FONCTIONS ALÉATOIRES STATIONNAIRES

17.1. Notion de processus aléatoire stationnaire

Dans la pratique on rencontre souvent des processus aléatoires se déroulant dans le temps d'une manière à peu près homogène et ressemblant à des oscillations aléatoires continues autour d'une certaine valeur moyenne, ni l'amplitude moyenne, ni le caractère de ces oscillations ne subissant de variations importantes dans le temps. Ces processus aléatoires sont dits *stationnaires*.

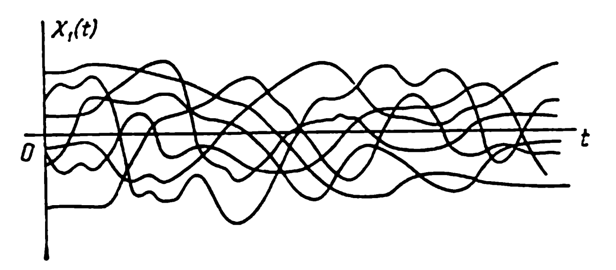


Fig. 17.1.1

A titre d'exemple de processus aléatoires stationnaires on peut citer : 1) les vibrations d'un avion en régime stable de vol horizontal; 2) les oscillations de la tension dans un réseau électrique ; 3) les bruits aléatoires dans un récepteur radio ; 4) le ballottement d'un navire, etc.

Un processus stationnaire peut être considéré comme se déroulant infiniment longtemps; lors de l'étude d'un processus stationnaire le choix du moment initial importe peu. Les caractéristiques d'un processus stationnaire doivent être les mêmes, quel que soit l'intervalle de temps qu'on choisit pour étudier le processus. D'une façon imagée on peut dire qu'un processus aléatoire « n'a ni commencement ni fin ».

Par exemple, la variation de la hauteur du centre de gravité d'un avion en régime stable de vol horizontal est un processus aléatoire stationnaire (fig. 17.1.1).

A l'opposé des processus stationnaires on peut citer des processus nettement non stationnaires, par exemple, les vibrations d'un avion en régime de plongée; les oscillations amorties dans un circuit électrique; la combustion de la charge propulsive dans une chambre de propulsion, etc. Un processus non stationnaire accuse toujours une certaine tendance de développement dans le temps; les caractéristiques d'un tel processus dépendent de l'origine et du temps.

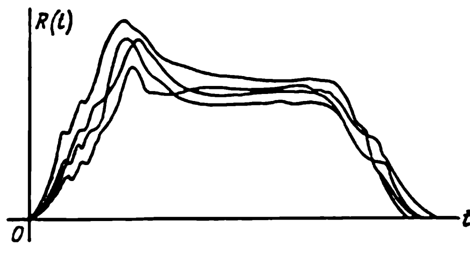


Fig. 17.1.2

Sur la figure 17.1.2 on peut voir une famille de réalisations d'un processus aléatoire manifestement non stationnaire, il s'agit du processus de variation dans le temps de la poussée d'un moteur.

Il y a lieu de noter qu'il existe des processus aléatoires manifestement non stationnaires qui ne le sont pas durant le cours de leur développement. Dans certains intervalles de temps ou avec une certaine approximation ils peuvent être considérés comme stationnaires.

En général tout processus aléatoire dans tout système dynamique commence par une phase non stationnaire appelée « processus transitoire ». Après l'amortissement du processus transitoire le système passe en général à un régime stable, les processus aléatoires s'y déroulant alors peuvent être considérés comme stationnaires.

Dans les applications physiques et techniques on rencontre très souvent des processus aléatoires stationnaires. De par leur nature ces processus sont plus simples que les processus non stationnaires et ainsi plus simples sont les caractéristiques les décrivant. Les transformations linéaires des processus aléatoires stationnaires sont également plus faciles à réaliser que dans le cas des processus non stationnaires. C'est pourquoi très employée en pratique est la théorie spéciale des *processus aléatoires stationnaires* ou plus exactement la *théorie des fonctions aléatoires stationnaires*, car l'argument d'une fonction aléatoire peut dans le cas général être autre que le temps. Le présent chapitre est consacré à un bref exposé de cette théorie.

Une fonction aléatoire $X(t)$ est dite *stationnaire* si toutes ses propriétés statistiques sont invariantes dans tout changement de l'origine des temps.

Dans notre exposé élémentaire de la théorie des fonctions aléatoires nous ne faisons pas appel à des lois de probabilité, en nous limitant uniquement à l'espérance mathématique, à la variance et à

la fonction de corrélation. Nous allons donc définir une fonction aléatoire stationnaire en nous basant sur ces caractéristiques.

Une fonction aléatoire stationnaire devant varier d'une manière homogène dans le temps, il est tout naturel de demander que l'espérance mathématique d'une fonction aléatoire stationnaire soit constante

$$m_x(t) = m_x = \text{const.} \tag{17.1.1}$$

Il y a lieu cependant de noter que cette condition n'est pas essentielle; en effet, on peut toujours passer de la fonction aléatoire $X(t)$ à la fonction aléatoire centrée $\overset{\circ}{X}(t)$ dont l'espérance mathématique est identiquement nulle et par conséquent satisfait à la condition

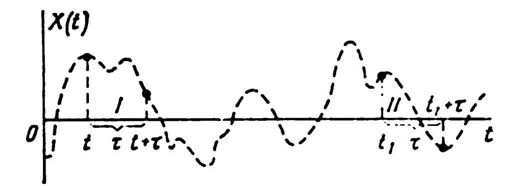


Fig. 17.1.3

(17.1.1). Ainsi, si un processus aléatoire n'est pas stationnaire que parce que son espérance mathématique est variable, ceci ne nous empêchera pas de l'étudier comme un processus stationnaire.

Une fonction aléatoire stationnaire doit également satisfaire à une seconde condition, à savoir sa variance doit être constante:

$$D_x(t) = D_x = \text{const.} \tag{17.1.2}$$

Essayons d'établir la condition à remplir par la fonction de corrélation d'une fonction aléatoire stationnaire. Soit une fonction aléatoire $X(t)$ (fig. 17.1.3). Dans l'expression de $K_x(t, t')$ posons $t' = t + \tau$ et soit $K_x(t, t+\tau)$ la covariance de deux sections de la fonction aléatoire séparées par l'intervalle de temps τ. Il est évident que si le processus aléatoire $X(t)$ est réellement stationnaire, cette covariance ne doit pas dépendre de l'origine des temps mais seulement de la longueur de cet intervalle. Par exemple, pour les domaines I et II sur la figure 17.1.3 de même longueur, les valeurs $K_x(t, t+\tau)$ et $K_x(t_1, t_1+\tau)$ de la fonction de corrélation doivent être les mêmes. D'une façon générale, la fonction de corrélation d'un processus aléatoire stationnaire doit satisfaire à:

$$K_x(t, t+\tau) = k_x(\tau). \tag{17.1.3}$$

Ceci signifie que la fonction de corrélation d'un processus aléatoire stationnaire est une fonction non pas de deux mais d'un seul argu-

ment, ce qui simplifie beaucoup les opérations sur les fonctions aléatoires stationnaires.

Il faut noter que la condition (17.1.2) exigeant que la variance d'une fonction aléatoire stationnaire soit constante est un cas particulier de (17.1.3). En effet, en posant dans la formule (17.1.3) $t + \tau = t$ ($\tau = 0$) on a

$$D_x(t) = K_x(t, t) = k_x(0) = \text{const.} \tag{17.1.4}$$

Ainsi, (17.1.3) est l'unique condition essentielle imposée à une fonction aléatoire stationnaire.

Dans l'exposé ultérieur nous entendrons par fonction aléatoire stationnaire une fonction aléatoire telle que sa fonction de corrélation dépende non pas de deux arguments t et t' mais seulement de

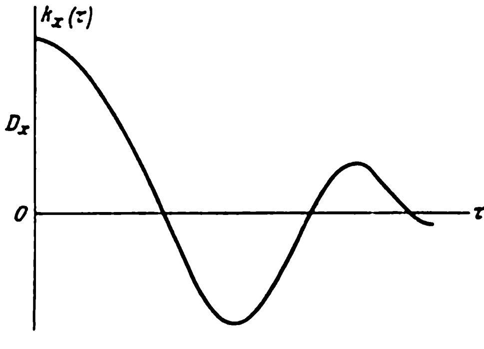


Fig. 17.1.4

l'intervalle les séparant. Pour ne pas imposer de conditions spéciales à l'espérance mathématique, nous n'envisagerons que des fonctions aléatoires centrées.

On sait que la fonction de corrélation d'une fonction aléatoire est symétrique :

$$K_x(t, t') = K_x(t', t).$$

Posant $t' - t = \tau$ on obtient pour un processus stationnaire :

$$k_x(\tau) = k_x(-\tau), \tag{17.1.5}$$

c'est-à-dire que la fonction de corrélation $k_x(\tau)$ est une *fonction paire*. C'est pourquoi on se borne en général à définir la fonction de corrélation pour les valeurs positives de son argument (fig. 17.1.4).

Dans la pratique on utilise couramment au lieu de la fonction de corrélation $k_x(\tau)$ la *fonction de corrélation normée* :

$$\rho_x(\tau) = \frac{k_x(\tau)}{D_x}, \tag{17.1.6}$$

où $D_x = k_x(0)$ est la variance constante du processus stationnaire. La fonction $\rho_x(\tau)$ n'est rien d'autre que le coefficient de corrélation entre les sections de la fonction aléatoire, séparées par un laps de temps égal à τ. Il est évident que l'on a $\rho_x(0) = 1$.

A titre d'exemple nous considérons deux échantillons de processus aléatoires approximativement stationnaires et nous déterminons leurs caractéristiques.

Exemple 1. La fonction aléatoire $X(t)$ est donnée par un ensemble de 12 réalisations (fig. 17.1.5). On demande de trouver: a) ses caractéristiques $m_x(t)$, $K_x(t, t')$, $D_x(t)$ et la fonction de corrélation normée $r_x(t, t')$; b) les mêmes caractéristiques en supposant $X(t)$ approximativement stationnaire.

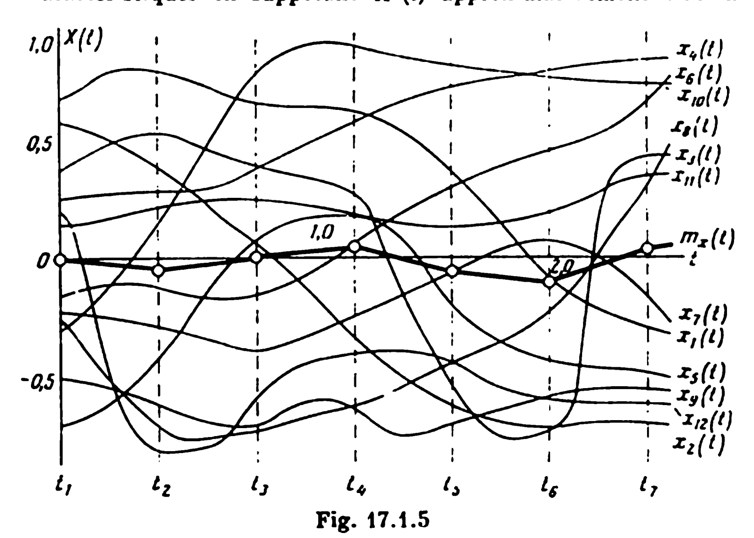


Fig. 17.1.5

Solution. Comme la fonction aléatoire $X(t)$ varie d'une manière relativement continue, les sections peuvent être relativement espacées, par exemple toutes les 0,4 s. Dans ce cas la fonction aléatoire se réduira à un système de sept variables aléatoires correspondant aux sections $t = 0$; 0,4; 0,8; 1,2; 1,6; 2,0; 2,4. En marquant ces sections sur le graphique on lit les valeurs de $X(t)$ dans ces sections et on établit le tableau des valeurs relevées (tableau 17.1.1).

Il est recommandé de remplir ce tableau ligne par ligne en se déplaçant tout le temps le long d'une même réalisation.

Puis on trouve les estimations des caractéristiques des variables aléatoires $X(0)$, $X(0,4)$, . . ., $X(2,4)$. En prenant la somme des colonnes et en divisant cette somme par le nombre de réalisations $n = 12$ on trouve approximativement l'espérance mathématique en fonction de temps:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{m}_x(t)$	−0,007	−0,057	0,000	0,037	−0,057	−0,093	0,036

Sur le graphique de la fig. 17.1.5 l'espérance mathématique est représentée en trait gras.

Tableau 17.1.1

t / n° de réalisation	0	0,4	0,8	1,2	1,6	2,0	2,4
1	0,64	0,74	0,62	0,59	0,35	−0,09	−0,39
2	0,54	0,37	0,06	−0,32	−0,60	−0,69	−0,67
3	0,34	0,50	0,37	0,26	−0,52	−0,72	0,42
4	0,23	0,26	0,35	0,55	0,69	0,75	0,80
5	0,12	0,20	0,24	0,18	−0,20	−0,42	−0,46
6	−0,16	−0,12	−0,15	0,05	0,29	0,43	0,63
7	−0,22	−0,29	−0,38	−0,24	−0,06	0,07	−0,16
8	−0,26	−0,69	−0,70	−0,61	−0,43	−0,22	0,29
9	−0,50	−0,60	−0,68	−0,62	−0,68	−0,56	−0,54
10	−0,30	0,13	0,75	0,84	0,78	0,73	0,71
11	−0,69	−0,40	0,08	0,16	0,12	0,18	0,33
12	0,18	−0,79	−0,56	−0,39	−0,42	−0,58	−0,53

Ensuite on trouve les estimations des éléments de la matrice de corrélation, à savoir des variances et des covariances. Le plus simple est de procéder comme suit. Pour le calcul de la variance statistique on prend la somme des carrés des nombres par colonne, on divise cette somme par $n = 12$ et on retranche du résultat le carré de l'espérance mathématique correspondante. Pour trouver l'estimation centrée le résultat doit être multiplié par le terme correcteur $\frac{n}{n-1} = \frac{12}{11}$. Les covariances sont estimées d'une manière analogue. Pour le calcul du moment statistique correspondant à deux sections données il y a lieu de multiplier entre eux les nombres se trouvant dans les colonnes respectives, de faire la somme algébrique des produits, de diviser par $n = 12$ la somme obtenue et de retrancher du résultat le produit des espérances mathématiques correspondantes. L'estimation centrée de la covariance s'obtient en multipliant le résultat par $\frac{n}{n-1}$. Lorsque les calculs sont effectués à l'aide d'une machine à calculer, les produits intermédiaires ne sont pas portés dans le tableau, mais s'ajoutent au total accumulé *). La matrice de corrélation du système des variables aléatoires $X(0)$, $X(0{,}4)$, . . ., $X(2{,}4)$ est donnée dans le tableau 17.1.2. C'est également le tableau des valeurs de la fonction de corrélation $\tilde{K}_x(t, t')$.

Suivant la diagonale principale on trouve les estimations des variances:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{D}_x(t)$	0,1632	0,2385	0,2356	0,2207	0,2407	0,2691	0,2878

*) Dans le cas présent il s'est trouvé possible d'effectuer le traitement directement à l'aide des moments initiaux, car l'espérance mathématique de la fonction $X(t)$ est voisine de zéro. Dans le cas contraire il y a lieu, avant de commencer le traitement, de transférer l'origine dans le voisinage de l'espérance mathématique.

Tableau 17.1.2

t' \ t	0	0,4	0,8	1,2	1,6	2,0	2,4
0	0,1632	0,1379	0,0795	0,0457	−0,0106	−0,0642	−0,0648
0,4		0,2385	0,2022	0,1621	0,0827	0,0229	0,0251
0,8			0,2356	0,2152	0,1527	0,0982	0,0896
1,2				0,2207	0,1910	0,1491	0,1322
1,6					0,2407	0,2348	0,1711
2,0						0,2691	0,2114
2,4							0,2878

En prenant la racine carrée de ces grandeurs on trouve l'écart quadratique moyen $\tilde{\sigma}_x$ en fonction du temps:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{\sigma}_x(t)$	0,404	0,488	0,485	0,470	0,491	0,519	0,536

En divisant les valeurs se trouvant dans le tableau 17.1.2 par les produits des écarts quadratiques moyens correspondants, on obtient le tableau des valeurs de la fonction de corrélation normée $\tilde{r}_x(t, t')$ (tableau 17.1.3).

Tableau 17.1.3

t' \ t	0	0,4	0,8	1,2	1,6	2,0	2,4
0	1	0,700	0,405	0,241	−0,053	−0,306	−0,299
0,4		1	0,856	0,707	0,345	0,090	0,095
0,8			1	0,943	0,643	0,390	0,344
1,2				1	0,829	0,612	0,524
1,6					1	0,923	0,650
2,0						1	0,760
2,4							1

Nous allons voir à quel point la fonction aléatoire $X(t)$ est stationnaire. A première vue les résultats de traitement ne permettent pas de conclure que la fonction aléatoire $X(t)$ est stationnaire. En effet, son espérance mathématique n'est pas rigoureusement constante, sa variance change également un peu dans le temps, les valeurs de la fonction de corrélation normée suivant les parallèles à la diagonale principale ne sont pas constantes non plus. Mais vu que le nombre de réalisations traitées n'est pas important ($n = 12$) et que les estimations obtenues dépendent donc fortement du hasard, on ne doit pas considérer les écarts observés comme importants, surtout qu'ils n'accusent pas de régularité. C'est pourquoi il est raisonnable de remplacer la fonction aléatoire $X(t)$ par une fonction stationnaire. Pour trouver cette fonction aléatoire stationnaire approchée, nous allons calculer la moyenne temporelle de l'estimation de l'es-

pérance mathématique:

$$\tilde{m}_x = \frac{\tilde{m}_x(0) + \tilde{m}_x(0,4) + \ldots + \tilde{m}_x(2,4)}{7} \approx -0,02.$$

D'une manière analogue on prend la moyenne de la variance:

$$\tilde{D}_x = \frac{\tilde{D}_x(0) + \tilde{D}_x(0,4) + \ldots + \tilde{D}_x(2,4)}{7} \approx 0,236.$$

En extrayant la racine carrée on trouve la moyenne de l'estimation de l'écart quadratique moyen:

$$\tilde{\sigma}_x \approx 0,486.$$

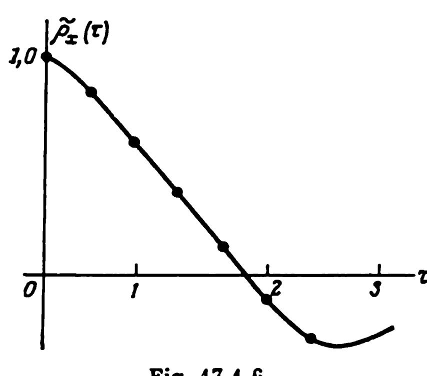


Fig. 17.1.6

Cherchons maintenant la fonction de corrélation normée du processus stationnaire pouvant servir d'approximation à la fonction $X(t)$. Pour un processus stationnaire la fonction de corrélation (donc également la fonction de corrélation normée) ne dépend que de $\tau = t' - t$; par conséquent, pour τ constant la fonction de corrélation doit être constante. Selon le tableau 17.1.3 à la condition $\tau = \text{const}$ satisfont la diagonale principale ($\tau = 0$) et les parallèles à cette diagonale ($\tau = 0,4$; $\tau = 0,8$; $\tau = 1,2$, etc.). En prenant la moyenne des estimations de la fonction de corrélation le long de ces parallèles à la diagonale principale on obtient la valeur de la fonction $\tilde{\rho}_x(\tau)$:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{\rho}_x(\tau)$	1,00	0,84	0,60	0,38	0,13	−0,10	−0,30

Le graphique de la fonction $\tilde{\rho}_x(\tau)$ est donné sur la figure 17.1.6.

En l'examinant on remarque que pour certaines valeurs de τ la fonction de corrélation est négative. Ceci traduit une certaine périodicité dans la structure de la fonction aléatoire, entraînant sur l'intervalle égal à peu près à la moitié de la période des oscillations fondamentales une corrélation négative entre les valeurs de la fonction aléatoire: à des écarts positifs par rapport à la moyenne dans une section correspondent des écarts négatifs dans une autre section distante de la première d'un certain intervalle de temps, et inversement.

Dans la pratique on rencontre assez souvent des fonctions de corrélation à valeurs négatives. Généralement dans ces cas, au fur et à mesure de l'augmentation de τ, l'amplitude des oscillations de la fonction de corrélation diminue et pour des τ grands la fonction de corrélation tend vers zéro.

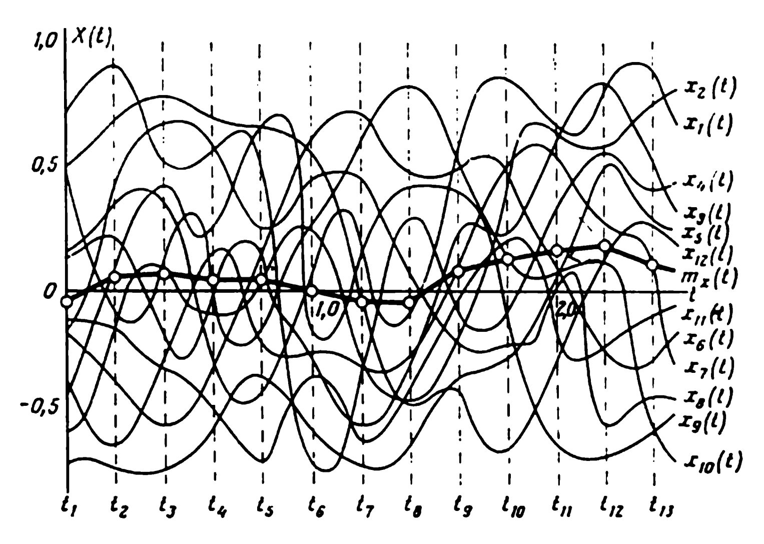


Fig. 17.1.7

Exemple 2. La fonction aléatoire $X(t)$ est donneé par un ensemble de 12 réalisations (fig. 17.1.7). En approchant la fonction $X(t)$ par une fonction stationnaire comparer la fonction de corrélation normée de cette dernière avec la fonction $\tilde{\rho}_x(\tau)$ de l'exemple précédent.

Solution. Comme la fonction aléatoire $X(t)$ varie d'une manière bien moins continue que la fonction $X(t)$ de l'exemple précédent, l'intervalle entre les sections ne peut plus être choisi égal à 0,4 s comme dans l'exemple précédent, mais au moins deux fois plus petit (par exemple, 0,2 s comme sur la figure 17.1.7). Le dépouillement des données fournit l'estimation de la fonction de corrélation normeé $\tilde{\rho}_x(\tau)$:

τ	0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0	2,2	2,4
$\tilde{\rho}_x(\tau)$	1,00	0,73	0,41	0,22	−0,01	−0,20	−0,19	−0,10	−0,06	−0,15	0,08	0,19	0,05

Sur la figure 17.1.8 on a représenté le graphique de la fonction $\tilde{\rho}_x(t)$. En comparant les courbes des figures 17.1.8 et 17.1.6 on voit que celle de la figure 17.1.8 décroît plus rapidement. Ce résultat est tout naturel, car l'allure de la fonction $X(t)$ dans l'exemple 1 est bien plus douce que dans l'exemple 2; par conséquent, la corrélation entre les valeurs de la fonction aléatoire de l'exemple 1 décroît plus lentement.

Ce qui frappe dans la courbe de la figure 17.1.8 c'est des oscillations inattendues de la fonction $\tilde{\rho}_x(\tau)$ pour des τ grands. Mais remarquant que pour des valeurs importantes de τ les points de la courbe sont obtenus à partir d'un

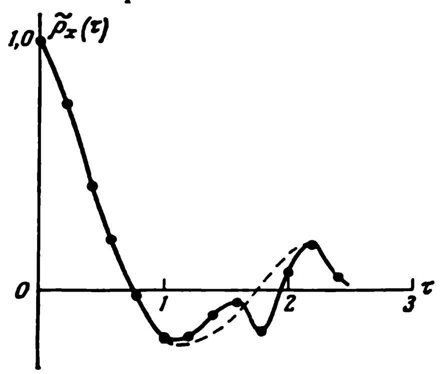


Fig. 17.1.8

nombre relativement restreint de données, on doit les considérer comme approximatifs. Dans ce cas il semble bon de procéder à un lissage de la fonction de corrélation comme ceci est montré en pointillé sur la figure 17.1.8.

17.2. Développement spectral d'une fonction aléatoire stationnaire sur un intervalle de temps fini. Spectre des variances

Les deux exemples donnés dans le paragraphe précédent montrent qu'il existe un lien entre la structure interne du processus aléatoire et l'allure de la fonction de corrélation correspondante, cette dernière dépendant fortement de la composition et des fréquences de la fonction aléatoire. Ceci nous amène directement à la notion de *composition spectrale* d'une fonction aléatoire.

La notion de « spectre » apparaît non seulement dans la théorie des fonctions aléatoires ; on la rencontre également en mathématiques, en physique et en technique.

Si un certain processus oscillatoire se présente comme la somme d'oscillations harmoniques de différentes fréquences (d'harmoniques), son *spectre* est la fonction donnant la répartition des amplitudes aux différentes fréquences. Le spectre indique quelles sont les oscillations qui dominent dans le processus étudié et quelle est sa structure interne.

De même on peut parler du spectre d'un processus aléatoire stationnaire ; seulement les amplitudes des oscillations seront des variables aléatoires. Le spectre d'une fonction aléatoire stationnaire décrira la *répartition des variances* suivant les fréquences.

Nous allons maintenant aborder la notion de spectre d'une fonction aléatoire stationnaire à partir des raisonnements suivants.

Considérons la fonction aléatoire stationnaire $\overset{\circ}{X}(t)$ observée sur l'intervalle (0, T) (fig. 17.2.1) et soit la fonction de corrélation

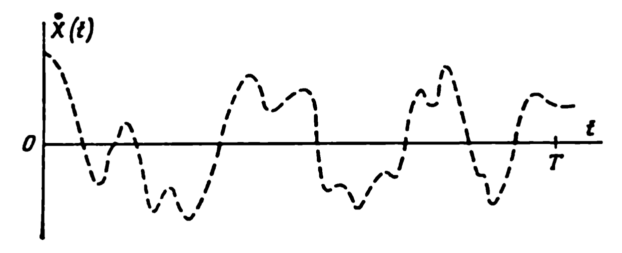


Fig. 17.2.1

de la fonction aléatoire $X(t)$:

$$K_x(t, t+\tau) = k_x(\tau).$$

La fonction $k_x(\tau)$ étant une fonction paire:

$$k_x(\tau) = k_x(-\tau),$$

sa courbe est symétrique (fig. 17.2.2).

Lorsque t et t' varient de 0 à T, l'argument $\tau = t' - t$ varie de $-T$ à $+T$.

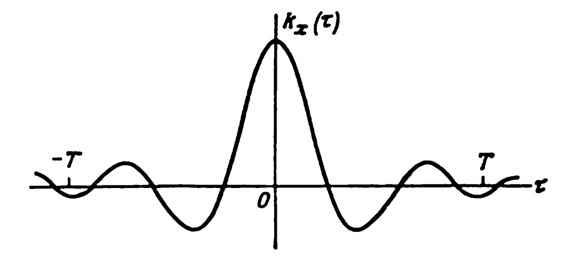


Fig. 17.2.2

Une fonction paire peut, sur l'intervalle $(-T, T)$, être développée en série de Fourier à l'aide des seuls harmoniques pairs (cosinus):

$$k_x(\tau) = \sum_{k=0}^{\infty} D_k \cos \omega_k \tau, \tag{17.2.1}$$

où

$$\omega_k = k\omega_1\,; \quad \omega_1 = \frac{2\pi}{2T} = \frac{\pi}{T}, \tag{17.2.2}$$

les coefficients D_x étant donnés par les formules:

$$\left.\begin{aligned} D_0 &= \frac{1}{2T}\int_{-T}^{T} k_x(\tau)\,d\tau,\\ D_k &= \frac{1}{T}\int_{-T}^{T} k_x(\tau)\cos\omega_k\tau\,d\tau \quad \text{pour } k\neq 0. \end{aligned}\right\} \tag{17.2.3}$$

Les fonctions $k_x(\tau)$ et $\cos\omega_k\tau$ étant les fonctions paires, on peut écrire comme suit les formules (17.2.3):

$$\left.\begin{aligned} D_0 &= \frac{1}{T}\int_{0}^{T} k_x(\tau)\,d\tau,\\ D_k &= \frac{2}{T}\int_{0}^{T} k_x(\tau)\cos\omega_k\tau\,d\tau \quad \text{pour } k\neq 0. \end{aligned}\right\} \tag{17.2.4}$$

Passons dans l'expression (17.2.1) de la fonction de corrélation $k_x(\tau)$ de l'argument τ à deux arguments t et t'. A cet effet posons:

$$\cos\omega_k\tau = \cos\omega_k(t'-t) = \cos\omega_k t'\cos\omega_k t + \sin\omega_k t'\sin\omega_k t \tag{17.2.5}$$

et portons (17.2.5) dans la formule (17.2.1):

$$K_x(t, t') = \sum_{k=0}^{\infty}(D_k\cos\omega_k t'\cos\omega_k t + D_k\sin\omega_k t'\sin\omega_k t). \tag{17.2.6}$$

On voit que l'expression (17.2.6) n'est rien d'autre que le *développement canonique* de la fonction de corrélation $K_x(t, t')$. Les fonctions de coordonnées de ce développement canonique sont alternativement les cosinus et les sinus des fréquences multiples de ω_1:

$$\cos\omega_k t, \quad \sin\omega_k t \quad (k = 0, 1, \ldots).$$

Nous avons déjà vu qu'il était possible de trouver à partir du développement canonique d'une fonction de corrélation le développement de la fonction aléatoire correspondante avec les mêmes fonctions de coordonnées et des variances égales aux coefficients D_k dans le développement canonique de la fonction de corrélation *).

Par conséquent, le développement canonique de la fonction aléatoire $\mathring{X}(t)$ s'écrit:

$$\mathring{X}(t) = \sum_{k=0}^{\infty}(U_k\cos\omega_k t + V_k\sin\omega_k t), \tag{17.2.7}$$

*) On peut montrer que pour une fonction de corrélation quelconque $k_x(\tau)$ les coefficients D_k donnés par les formules (17.2.4) ne sont pas négatifs.

où U_k, V_k sont des variables aléatoires non corrélées d'espérances mathématiques nulles et d'égales variances pour tout couple de variables de même indice k:

$$D[U_k] = D[V_k] = D_k. \qquad (17.2.8)$$

Les variances pour différentes valeurs de k sont données par les formules (17.2.4).

Nous avons ainsi obtenu sur l'intervalle $(0, T)$ le développement canonique de la fonction aléatoire $\overset{\circ}{X}(t)$ dont les fonctions de coordonnées sont $\cos \omega_k t$, $\sin \omega_k t$ pour différents ω_k. Ce développement est

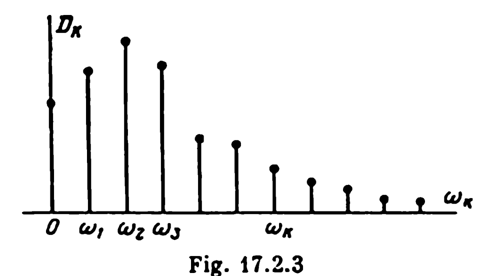


Fig. 17.2.3

appelé *développement spectral* d'une fonction aléatoire stationnaire. Il est à la base de la *théorie spectrale* des processus aléatoires stationnaires.

Le développement spectral représente une fonction aléatoire stationnaire comme décomposée suivant les oscillations harmoniques de différentes fréquences

$$\omega_1, \omega_2, \ldots, \omega_k, \ldots$$

d'amplitudes aléatoires.

Calculons la variance de la fonction aléatoire $\overset{\circ}{X}(t)$ donnée par le développement spectral (17.2.7). En vertu du théorème sur la variance d'une fonction linéaire de variables aléatoires non corrélées on a:

$$D_x = D[\overset{\circ}{X}(t)] = \sum_{k=0}^{\infty} (\cos^2 \omega_k t + \sin^2 \omega_k t) D_k = \sum_{k=0}^{\infty} D_k. \qquad (17.2.9)$$

Ainsi, la *variance d'une fonction aléatoire stationnaire est égale à la somme des variances de tous les harmoniques de son développement spectral*. La formule (17.2.9) montre que la variance de la fonction $\overset{\circ}{X}(t)$ possède une certaine répartition suivant différentes fréquences: à certaines fréquences correspondent des variances plus importantes, à d'autres moins importantes. Graphiquement, on peut représenter cette répartition comme le spectre d'une fonction aléatoire stationnaire (plus exactement, comme le spectre des variances). A cet effet

on porte sur l'axe des abscisses les fréquences $\omega_0 = 0$, ω_1, ω_2, ..., ω_k, ... et suivant l'axe des ordonnées les variances correspondantes (fig. 17.2.3).

Il est évident que la somme de toutes les ordonnées du spectre est égale à la variance de la fonction aléatoire.

17.3. Développement spectral d'une fonction aléatoire stationnaire sur un intervalle de temps infini. Densité spectrale d'une fonction aléatoire stationnaire

Lorsqu'on procède à la décomposition spectrale d'une fonction aléatoire stationnaire $\mathring{X}(t)$ sur un intervalle de temps fini $(0, T)$, on obtient le spectre des variances de cette fonction sous la forme d'une série de lignes discrètes séparées par des intervalles de temps égaux (spectre discontinu ou de raies).

Il est évident que plus grand est l'intervalle de temps étudié, plus complètes sont nos données sur la fonction aléatoire. Ainsi, dans le développement spectral il est naturel d'essayer de passer à la limite pour $T \to \infty$ pour voir ce que devient dans ce cas le spectre de la fonction aléatoire. Pour $T \to \infty$, $\omega_1 = \frac{2\pi}{2T} \to 0$; ainsi, la distance séparant les fréquences ω_k du spectre diminuera indéfiniment. Le spectre discret tendra vers le spectre continu dans lequel à tout petit intervalle de fréquences $\Delta\omega$ correspondra la variance élémentaire $\Delta D(\omega)$.

Essayons de représenter graphiquement un spectre continu. A cet effet, il faut quelque peu modifier le graphique du spectre discret pour T fini. Nous porterons sur l'axe des ordonnées non plus la variance D_k (diminuant indéfiniment pour $T \to \infty$), mais la *densité moyenne de variance*, c'est-à-dire la variance rapportée à l'unité de longueur de l'intervalle de fréquences correspondant. Désignons par $\Delta\omega$ la distance entre les fréquences voisines:

$$\omega_1 = \frac{2\pi}{2T} = \Delta\omega.$$

Prenant pour base le segment $\Delta\omega$ construisons le rectangle d'aire D_k (fig. 17.3.1). Nous obtenons ainsi un diagramme en escalier rappelant l'histogramme de la répartition statistique.

La hauteur du rectangle sur l'intervalle $\Delta\omega$ attenant au point ω_k est égale à

$$S_x(\omega_k) = \frac{D_k}{\Delta\omega}, \tag{17.3.1}$$

elle correspond à la densité moyenne de la variance sur cet intervalle. L'aire totale de tout le diagramme est de toute évidence égale à la variance de la fonction aléatoire.

Tendons vers l'infini l'intervalle T. On aura alors $\Delta\omega \to 0$ et la courbe en escalier se rapprochera à la limite de la courbe continue $S_x(\omega)$ (fig. 17.3.2). Cette courbe représente la densité de la répartition des probabilités de la variance suivant les fréquences du spectre

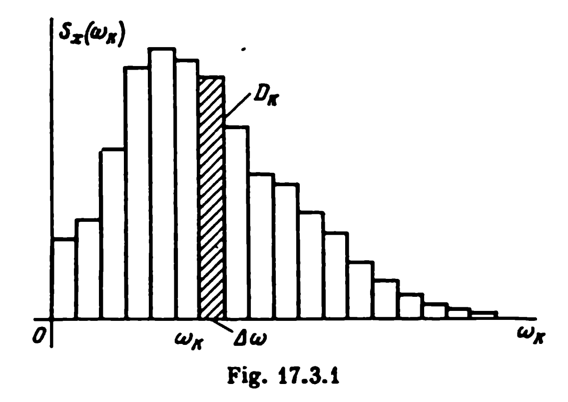


Fig. 17.3.1

continu, la fonction $S_x(\omega)$ portant le nom de *densité spectrale de la variance* ou plus brièvement *densité spectrale* de la fonction aléatoire stationnaire $\overset{\circ}{X}(t)$.

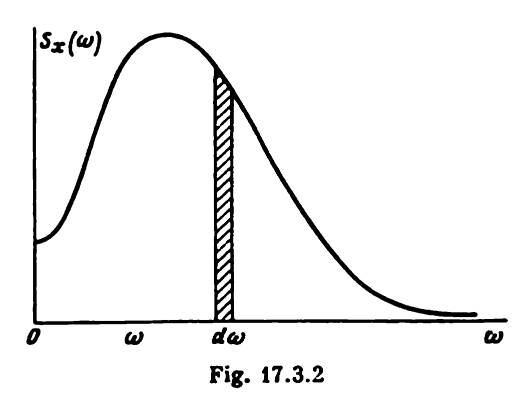


Fig. 17.3.2

Il est évident que, comme auparavant, la surface limitée par la courbe $S_x(\omega)$ doit être égale à la variance D_x de la fonction aléatoire $\overset{\circ}{X}(t)$:

$$D_x = \int_0^\infty S_x(\omega)\,d\omega. \qquad (17.3.2)$$

La formule (17.3.2) n'est rien d'autre que le développement de la variance D_x en une somme de composantes élémentaires $S_x(\omega)\,d\omega$

égales chacune à la variance correspondant à l'intervalle élémentaire de fréquences $d\omega$ attenant au point ω (fig. 17.3.2).

Nous avons ainsi introduit une nouvelle caractéristique d'un processus aléatoire stationnaire, à savoir la densité spectrale, donnant la composition spectrale du processus stationnaire. Mais cette caractéristique n'est pas indépendante, étant entièrement déterminée par la fonction de corrélation du processus étudié. De même que les ordonnées D_k du spectre discret s'expriment à l'aide des formules (17.2.4) par la fonction de corrélation $k_x(\tau)$, la densité spectrale $S_x(\omega)$ peut être exprimée par la fonction de corrélation.

Nous allons trouver cette expression. A cet effet nous passons dans le développement canonique de la fonction de corrélation à la limite pour $T \to \infty$ et voyons ce qu'il devient. Le développement (17.2.1) de la fonction de corrélation en série de Fourier sur l'intervalle fini $(-T, T)$ s'écrit:

$$k_x(\tau) = \sum_{k=0}^{\infty} D_k \cos \omega_k \tau, \tag{17.3.3}$$

la variance correspondant à la fréquence ω_k étant égale à:

$$D_k = \frac{2}{T} \int_0^T k_x(\tau) \cos \omega_k \tau \, d\tau. \tag{17.3.4}$$

Avant de passer à la limite pour $T \to \infty$ passons dans la formule (17.3.3) de la variance D_k à la densité moyenne de la variance $\frac{D_k}{\Delta\omega}$. Calculée pour une valeur finie de T cette densité dépend de T, nous la désignerons:

$$S_x^{(T)}(\omega_k) = \frac{D_k}{\Delta\omega}. \tag{17.3.5}$$

En divisant l'expression (17.3.4) par $\Delta\omega = \frac{\pi}{T}$, nous obtenons:

$$S_x^{(T)}(\omega_k) = \frac{2}{\pi} \int_0^T k_x(\tau) \cos \omega_k \tau \, d\tau. \tag{17.3.6}$$

En vertu de (17.3.5) on a:

$$D_k = S_x^{(T)}(\omega_k) \, \Delta\omega. \tag{17.3.7}$$

Le report de l'expression (17.3.7) dans la formule (17.3.3) donne:

$$k_x(\tau) = \sum_{k=0}^{\infty} S_x^{(T)}(\omega_k) \cos \omega_k \tau \Delta\omega. \tag{17.3.8}$$

Voyons ce que devient l'expression (17.3.8) pour $T \to \infty$. Il est clair qu'alors $\Delta\omega \to 0$; l'argument discret ω_k devient l'argument ω à variation continue; la somme devient une intégrale en ω;

la densité moyenne de la variance $S_x^{(T)}(\omega_k)$ tend vers la densité spectrale $S_x(\omega)$, l'expression (17.3.8) devient alors à la limite:

$$k_x(\tau) = \int_0^\infty S_x(\omega) \cos \omega\tau \, d\omega, \tag{17.3.9}$$

où $S_x(\omega)$ est la densité spectrale de la fonction aléatoire stationnaire.

En passant à la limite pour $T \to \infty$ dans la formule (17.3.6) on obtient l'expression de la densité spectrale en fonction de la fonction de corrélation:

$$S_x(\omega) = \frac{2}{\pi} \int_0^\infty k_x(\tau) \cos \omega\tau \, d\tau. \tag{17.3.10}$$

En mathématiques l'expression (17.3.9) porte le nom *d'intégrale de Fourier*. L'intégrale de Fourier est une généralisation du développement en série de Fourier dans le cas des fonctions non périodiques sur un intervalle de temps infini; c'est la décomposition d'une fonction en une somme d'oscillations harmoniques élémentaires à spectre continu *).

Tout comme la série de Fourier exprime la fonction développée en fonction des coefficients de la série, ceux-ci s'exprimant à leur tour en fonction de la fonction développée, les grandeurs $k_x(\tau)$ et $S_x(\omega)$ s'expriment l'une en fonction de l'autre. Les formules du type (17.3.9) et (17.3.10) reliant réciproquement deux fonctions sont appelées *transformations de Fourier* **).

Ainsi, la fonction de corrélation et la densité spectrale sont liées par des transformations de Fourier.

Notons que si dans la formule générale (17.3.9) on pose $\tau = 0$, on obtient l'expression (17.3.2) de la décomposition harmonique de la variance.

En pratique on utilise souvent au lieu de la densité spectrale $S_x(\omega)$ la densité spectrale *normée*:

$$s_x(\omega) = \frac{S_x(\omega)}{D_x}, \tag{17.3.11}$$

où D_x est la variance de la fonction aléatoire.

Il est facile de voir que la fonction de corrélation normée $\rho_x(\tau)$ et la densité spectrale normée $s_x(\omega)$ sont liées entre elles par les mêmes

---

*) La formule (17.3.9) est un cas particulier de l'intégrale de Fourier généralisant le développement en série de Fourier d'une fonction paire suivant les harmoniques cosinus. On peut écrire une expression analogue dans un cas plus général.

**) Les formules citées sont le cas particulier des transformations de Fourier, dites transformations cosinus.

transformations de Fourier:

$$\left.\begin{aligned} \rho_x(\tau) &= \int_0^\infty s_x(\omega)\cos\omega\tau\, d\omega, \\ s_x(\omega) &= \frac{2}{\pi}\int_0^\infty \rho_x(\tau)\cos\omega\tau\, d\tau. \end{aligned}\right\} \qquad (17.3.12)$$

En posant $\tau = 0$ dans la première égalité (17.3.12) et compte tenu de $\rho_x(0) = 1$, on a:

$$\int_0^\infty s_x(\omega)\, d\omega = 1, \qquad (17.3.13)$$

ce qui traduit le fait que l'aire limitée par la courbe de la densité spectrale normée est égale à l'unité.

Exemple 1. La fonction de corrélation normée $\rho_x(\tau)$ de la fonction aléatoire $X(t)$ décroît linéairement de l'unité à zéro pour $0 < \tau < \tau_0$; pour $\tau > \tau_0$ on a $\rho_x(\tau) = 0$ (fig. 17.3.3). Trouver la densité spectrale normée de la fonction aléatoire $X(t)$.

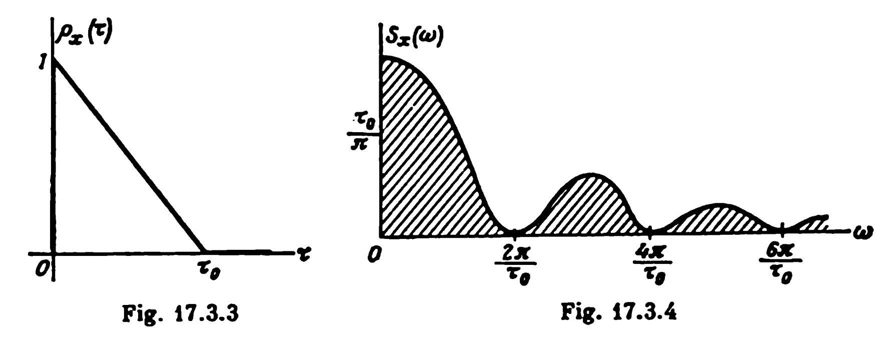


Fig. 17.3.3

Fig. 17.3.4

Solution. La fonction de corrélation normée est donnée par les formules:

$$\rho_x(\tau)\begin{cases} = 1 - \dfrac{\tau}{\tau_0} & \text{pour } 0 < \tau < \tau_0, \\ = 0 & \text{pour } \tau > \tau_0. \end{cases}$$

Les formules (17.3.12) donnent:

$$s_x(\omega) = \frac{2}{\pi}\int_0^\infty \rho_x(\tau)\cos\omega\tau\, d\tau = \frac{2}{\pi}\int_0^{\tau_0}\left(1 - \frac{\tau}{\tau_0}\right)\cos\omega\tau\, d\tau = \frac{2}{\pi\tau_0\omega^2}(1 - \cos\omega\tau_0).$$

Le graphique de la densité spectrale normée est donné sur la figure 17.3.4. Le premier maximum, absolu, de la densité spectrale correspond à $\omega = 0$; en levant l'indétermination en ce point on obtient $\frac{\tau_0}{\pi}$. Puis au fur et à mesure

de l'augmentation de ω la densité spectrale admet plusieurs maxima relatifs, décroissant avec l'augmentation de ω; pour $\omega \to \infty$ on a $s_x(\omega) \to 0$.

Le caractère des variations de la densité spectrale $s_x(\omega)$ (décroissance rapide ou lente) dépend du paramètre τ_0. L'aire totale limitée par la courbe $s_x(\omega)$ est constante et égale à l'unité. La variation de τ_0 équivaut à un changement de l'échelle de la courbe suivant les deux axes en conservant son aire. Lorsque τ_0 augmente, l'échelle s'étend en ordonnées et diminue en abscisses; la prédominance de la fréquence 0 dans le spectre de la fonction aléatoire devient plus nette. A la limite pour $\tau_0 \to \infty$ la fonction aléatoire devient une variable aléatoire ordinaire; on a alors $\rho_x(\tau) \equiv 1$ et le spectre devient discret avec une seule et unique raie de fréquence $\omega_0 = 0$.

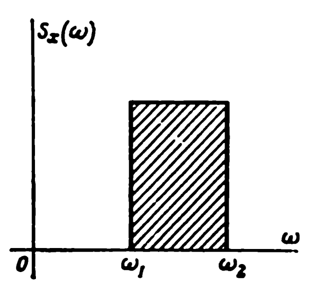


Fig. 17.3.5

E x e m p l e 2. La densité spectrale normée $s_x(\omega)$ de la fonction aléatoire $X(t)$ est constante sur un certain intervalle de fréquences ω_1, ω_2 et nulle à l'extérieur de cet intervalle (fig. 17.3.5).

Trouver la fonction de corrélation normée de la fonction aléatoire $X(t)$.

S o l u t i o n. La valeur de $s_x(\omega)$ pour $\omega_1 < \omega < \omega_2$ s'obtient à partir de la condition que l'aire limitée par la courbe $s_x(\omega)$ est égale à l'unité:

$$s_x(\omega)(\omega_2-\omega_1)=1, \quad s_x(\omega)=\frac{1}{\omega_2-\omega_1}.$$

En vertu de (17.3.12) on a:

$$\rho_x(\tau)=\int_{\omega_1}^{\omega_2} s_x(\omega)\cos\omega\tau\,d\omega=\frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2}\cos\omega\tau\,d\omega=$$

$$=\frac{1}{\tau(\omega_2-\omega_1)}(\sin\omega_2\tau-\sin\omega_1\tau)=\frac{2}{\tau(\omega_2-\omega_1)}\cos\left(\frac{\omega_2+\omega_1}{2}\tau\right)\sin\left(\frac{\omega_2-\omega_1}{2}\tau\right).$$

L'allure générale de la fonction $\rho_x(\tau)$ est montrée sur la figure 17.3.6. Ce sont des oscillations d'amplitude décroissante avec des nœuds là où la fonction s'annule. Dans chaque cas particulier la courbe dépend des valeurs de ω_1, ω_2.

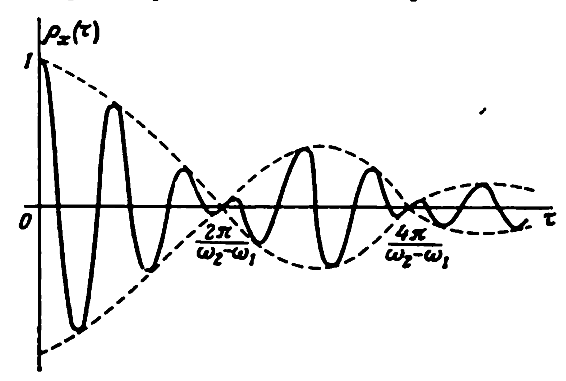


Fig. 17.3.6

Il est intéressant d'étudier le cas limite quand $\omega_1 \to \omega_2$. Il est évident que pour $\omega_2 = \omega_1 = \omega$ le spectre de la fonction aléatoire devient une seule

et unique raie apparaissant à la fréquence ω; la fonction de corrélation est alors une simple cosinusoïde:

$$\rho_x(\tau) = \cos \omega\tau.$$

Proposons-nous de déterminer l'allure de la fonction $\overset{\circ}{X}(t)$ dans ce cas. Le développement spectral de la fonction aléatoire stationnaire $\overset{\circ}{X}(t)$ dans le cas d'un spectre discret à raie unique sera:

$$\overset{\circ}{X}(t) = U \cos \omega t + V \sin \omega t, \qquad (17.3.14)$$

où U et V sont des fonctions aléatoires non corrélées d'espérances mathématiques nulles et de variances

$$D[U] = D[V] = D.$$

Montrons qu'une fonction aléatoire du type (17.3.14) peut être représentée par une oscillation harmonique de fréquence d'amplitude et de phase aléatoires. En introduisant les désignations:

$$\cos \Phi = \frac{U}{\sqrt{U^2+V^2}}; \quad \sin \Phi = \frac{V}{\sqrt{U^2+V^2}},$$

l'expression (17.3.14) s'écrit:

$$\overset{\circ}{X}(t) = \sqrt{U^2+V^2}(\cos \Phi \cos \omega t + \sin \Phi \sin \omega t) = \sqrt{U^2+V^2} \cos(\omega t - \Phi).$$

Dans cette expression $\sqrt{U^2+V^2}$ est l'amplitude aléatoire et Φ la phase aléatoire de l'oscillation harmonique.

Jusqu'à présent nous n'avons envisagé que le cas continu de répartition des variances lorsqu'à un intervalle de fréquences infiniment petit correspond une variance infiniment petite. Dans la pratique on rencontre des cas où la fonction aléatoire contient une composante purement périodique de fréquence ω_k et d'amplitude aléatoire. Alors le développement de la fonction aléatoire contient en plus du spectre continu de fréquences, une raie isolée à la fréquence ω_k de variance finie D_k. Dans le cas général le nombre de composantes périodiques peut être quelconque. Il est clair que le spectre de la fonction de corrélation se décomposera alors en deux parties: discrète et continue:

$$k_x(\tau) = \sum_k D_k \cos \omega_k \tau + \int_0^\infty S_x(\omega) \cos \omega\tau \, d\omega. \qquad (17.3.15)$$

Des cas de fonctions aléatoires stationnaires à spectre mixte sont relativement rares. Il y a toujours lieu de séparer les composantes à spectre continu de celles à spectre discret, et de les étudier séparément.

Assez souvent on rencontre le cas particulier où la variance finie dans le développement spectral de la fonction aléatoire correspond à la fréquence nulle ($\omega = 0$). Ceci traduit le fait que la fonction aléatoire contient comme composante une variable aléatoire ordinaire de variance D_0. Dans ce cas il y a lieu d'étudier séparément cette composante.

17.4. Développement spectral d'une fonction aléatoire sous forme complexe

Souvent les opérations sur les fonctions aléatoires se simplifient beaucoup si l'on écrit leur développement spectral, la densité spectrale et la fonction de corrélation non pas sous la forme réelle, mais complexe. En particulier, les opérations linéaires les plus diverses sur les fonctions ayant la forme d'oscillations harmoniques (dérivation, intégration, résolution des équations différentielles linéaires, etc.), sont bien plus simples lorsque ces oscillations sont écrites non pas à l'aide des sinus et des cosinus, mais comme la somme d'exponentielles complexes. L'emploi de la forme complexe de représentation de la fonction de corrélation et de la densité spectrale est justifié même si la fonction aléatoire (donc également sa fonction de corrélation et sa densité spectrale) est réelle.

Montrons comment on peut passer, formellement, de l'écriture réelle à la forme complexe de représentation du développement spectral d'une fonction aléatoire.

Soit le développement spectral (17.2.8) de la fonction aléatoire $\overset{\circ}{X}(t)$ sur l'intervalle $(0, T)$:

$$\overset{\circ}{X}(t) = \sum_{k=0}^{\infty} (U_k \cos \omega_k t + V_k \sin \omega_k t), \qquad (17.4.1)$$

où U_k, V_k sont des variables aléatoires non corrélées, d'égales variances pour chaque couple U_k, V_k de même indice valant:

$$D[U_k] = D[V_k] = D_k.$$

Comme $\omega_k = k\omega_1$ et $\omega_0 = 0$ l'expression (17.4.1) s'écrit:

$$\overset{\circ}{X}(t) = U_0 + \sum_{k=1}^{\infty} (U_k \cos \omega_k t + V_k \sin \omega_k t). \qquad (17.4.2)$$

Ecrivons maintenant le développement spectral (17.4.2) sous forme complexe. A cet effet il y a lieu d'utiliser les formules bien connues d'Euler:

$$\cos \omega_k t = \frac{e^{i\omega_k t} + e^{-i\omega_k t}}{2};$$

$$\sin \omega_k t = \frac{e^{i\omega_k t} - e^{-i\omega_k t}}{2i} = -i\,\frac{e^{i\omega_k t} - e^{-i\omega_k t}}{2}.$$

En substituant ces expressions dans la formule (17.4.2) on a:

$$\overset{\circ}{X}(t) = U_0 + \sum_{k=1}^{\infty} \left(U_k \frac{e^{i\omega_k t} + e^{-i\omega_k t}}{2} - iV_k \frac{e^{i\omega_k t} - e^{-i\omega_k t}}{2} \right), \qquad (17.4.3)$$

qui est le développement avec les fonctions de coordonnées $e^{i\omega_k t}$, $e^{-i\omega_k t}$.

Transformons le développement (17.4.3) de telle sorte que les fonctions de coordonnées soient uniquement $e^{i\omega_k t}$; à cet effet nous convenons d'étendre le domaine des fréquences ω_k aux valeurs négatives de ω et considérons donc les fréquences

$$\omega_k = k\omega_1 \quad (k = \pm 1, \pm 2, \ldots),$$

supposant que k prend non seulement des valeurs positives mais également négatives. La formule (17.4.3) s'écrira alors:

$$\overset{\circ}{X}(t) = U_0 + \sum_{k=1}^{\infty} \frac{U_k - iV_k}{2} e^{i\omega_k t} + \sum_{k=-1}^{-\infty} \frac{U_k + iV_k}{2} e^{i\omega_k t} \qquad (17.4.4)$$

en supposant que:

$$U_{-k} = U_k; \quad V_{-k} = V_k.$$

La formule (17.4.4) donne le développement de la fonction aléatoire $\overset{\circ}{X}(t)$, dans lequel les fonctions de coordonnées sont les fonctions complexes $e^{i\omega_k t}$ et les coefficients sont des variables aléatoires complexes. En désignant par W_k $(k = \pm 1, \pm 2, \ldots)$ ces variables aléatoires complexes, le développement (17.4.4) s'écrit:

$$\overset{\circ}{X}(t) = \sum_{k=-\infty}^{\infty} W_k e^{i\omega_k t}, \qquad (17.4.5)$$

où

$$\left.\begin{aligned} W_k &= U_0 && \text{pour } k = 0; \\ W_k &= \frac{U_k - iV_k}{2} && \text{pour } k > 0; \\ W_k &= \frac{U_k - iV_k}{2} && \text{pour } k < 0. \end{aligned}\right\} \qquad (17.4.6)$$

Nous allons maintenant montrer que l'expression (17.4.5) est bien le *développement canonique* de la fonction aléatoire $\overset{\circ}{X}(t)$. A cet effet il suffit de montrer que les coefficients aléatoires de ce développement ne sont pas intercorrélés.

Soient tout d'abord les coefficients W_k et W_l de deux termes différents dans la partie positive du spectre pour $k \neq l$, $k > 0$, $l > 0$; nous allons calculer le moment de corrélation de ces grandeurs. Par définition pour des variables aléatoires complexes on a:

$$K_{kl} = M[W_k \overline{W}_l],$$

où $\overline{W}_l$ est le complexe conjugué de W_l.

Pour $k > 0$, $l > 0$ on a

$$K_{kl} = M\left[\frac{U_k - iV_k}{2}\,\overline{\frac{U_l - iV_l}{2}}\right] = M\left[\frac{U_k - iV_k}{2}\,\frac{U_l + iV_l}{2}\right] =$$

$$= \frac{1}{4}\{M[U_kU_l] + iM[U_kV_l] - iM[U_lV_k] + M[V_kV_l]\} = 0,$$

car les variables aléatoires U_k, V_k figurant dans le développement (17.4.1) ne sont pas corrélées entre elles.

On montre d'une manière analogue que les variables aléatoires W_k, W_l ne sont pas corrélées quels que soient les signes des indices k et l si $k \neq \pm l$.

Il ne reste plus qu'à démontrer l'absence de corrélation entre les coefficients des termes symétriques du développement, c'est-à-dire les grandeurs W_k et W_{-k} pour k quelconque. On a :

$$K_{k,-k} = M[W_kW_{-k}] = M\left[\frac{U_k - iV_k}{2}\,\overline{\frac{U_k + iV_k}{2}}\right] =$$

$$= \frac{1}{4}M[(U_k - iV_k)^2] = \frac{1}{4}\{M[U_k^2] - M[V_k^2] - 2iM[U_kV_k]\}.$$

Comme les grandeurs U_k, V_k figurant dans un même terme du développement (17.4.1) ne sont pas corrélées et ont les variances égales D_k on obtient :

$$K_{k,-k} = \frac{1}{4}\{D_k - D_k - 2i\cdot 0\} = 0.$$

Nous avons ainsi montré que l'expression (17.4.5) est bien le développement canonique de la fonction $\overset{\circ}{X}(t)$ avec les fonctions de coordonnées complexes $e^{i\omega_k t}$ et les coefficients complexes W_k.

Calculons les variances de ces coefficients. Pour $k = 0$ la variance D_0 reste telle qu'elle était dans le développement spectral réel. La variance de chacune des variables complexes W_k (pour $k \neq 0$) est égale à la somme des variances de ses parties réelle et imaginaire :

$$D[W_k] = \frac{D[U_k]}{4} + \frac{D[V_k]}{4} = \frac{2D[U_k]}{4} = \frac{D_k}{2}.$$

Désignons

$$D_k^* = \frac{D_k}{2} \text{ pour } k \neq 0; \quad D_k^* = D_0 \text{ pour } k = 0$$

et construisons le spectre discret de la fonction aléatoire $\overset{\circ}{X}(t)$ étendu aux fréquences de $-\infty$ à $+\infty$ (fig. 17.4.1).

Ce spectre est symétrique par rapport à l'axe des ordonnées ; à la différence du spectre de la figure 17.2.3, il est déterminé non seulement pour les valeurs positives des fréquences mais également pour les valeurs négatives, mais les ordonnées pour $k \neq 0$ sont deux fois moindres que dans l'exemple précédent ; la somme de toutes les ordonnées reste, comme auparavant, égale à la variance de la fonc-

tion aléatoire $\dot{X}(t)$:

$$D_x = \sum_{k=-\infty}^{\infty} D_k. \tag{17.4.7}$$

Déterminons la fonction de corrélation de la fonction aléatoire $\dot{X}(t)$ écrite sous la forme d'un développement spectral complexe (17.4.5). En appliquant la formule (16.2.15) donnant la fonction de

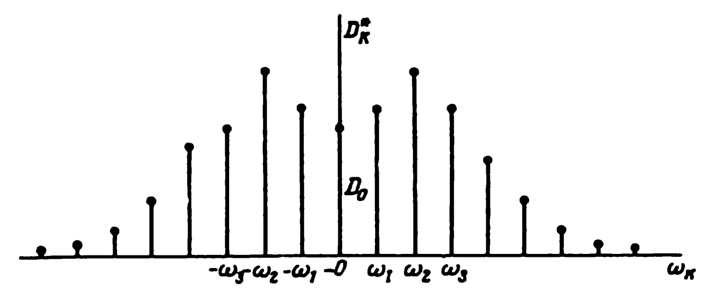


Fig. 17.4.1

corrélation d'une fonction aléatoire complexe écrite sous la forme d'un développement canonique on obtient:

$$k_x(t, t') = \sum_{k=-\infty}^{\infty} D_k^* e^{i\omega_k t'} \overline{e^{i\omega_k t}} = \sum_{k=-\infty}^{\infty} D_k^* e^{i\omega_k t'} e^{-i\omega_k t} = \sum_{k=-\infty}^{\infty} D_k^* e^{i\omega_k (t'-t)}$$

ou en passant à l'argument $\tau = t' - t$:

$$k_x(\tau) = \sum_{k=-\infty}^{\infty} D_k^* e^{j\omega_k \tau}, \tag{17.4.8}$$

avec

$$D_k^* = \frac{1}{2} D_k = \frac{1}{T} \int_0^T k_x(\tau) \cos \omega_k \tau \, d\tau \quad \text{pour } k \neq 0. \tag{17.4.9}$$

Ecrivons également (17.4.9) sous forme complexe. Posant:

$$\cos \omega_k \tau = \frac{e^{i\omega_k \tau} + e^{-i\omega_k \tau}}{2},$$

on obtient:

$$D_k^* = \frac{1}{2T} \int_0^T k_x(\tau) \left(e^{i\omega_k \tau} + e^{-i\omega_k \tau}\right) d\tau =$$

$$= \frac{1}{2T} \left\{ \int_0^T k_x(\tau) e^{-i\omega_k \tau} d\tau + \int_0^T k_x(\tau) e^{i\omega_k \tau} d\tau \right\}.$$

Posant dans la seconde intégrale $\tau = -u$ on a:

$$\int_0^T k_x(\tau)\, e^{i\omega_k \tau}\, d\tau = -\int_0^{-T} k_x(u)\, e^{-i\omega_k u}\, du = \int_{-T}^0 k_x(\tau)\, e^{-i\omega_k \tau}\, d\tau,$$

d'où

$$D_k^* = \frac{1}{2T} \int_{-T}^{T} k_x(\tau)\, e^{-i\omega_k \tau}\, d\tau. \qquad (17.4.10)$$

Nous avons ainsi obtenu la forme complexe du développement spectral d'une fonction aléatoire sur l'intervalle fini $(0, T)$. Il est naturel ensuite de passer à la limite pour $T \to \infty$ comme nous l'avons fait pour la forme réelle. En introduisant la densité spectrale:

$$S_x^*(\omega) = \lim_{\Delta\omega \to 0} \frac{D_k^*}{\Delta\omega}$$

nous obtenons, à la limite, à partir des formules (17.4.8), (17.4.10), les formules intégrales reliant la fonction de corrélation et la densité spectrale complexe. A la limite pour $T \to \infty$ les formules (17.4.8) et (17.4.10) donnent:

$$k_x(\tau) = \int_{-\infty}^{\infty} S_x^*(\omega)\, e^{i\omega\tau}\, d\omega, \qquad (17.4.11)$$

$$S_x^*(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k_x(\tau)\, e^{-i\omega\tau}\, d\tau. \qquad (17.4.12)$$

Les formules (17.4.11) et (17.4.12) sont la forme complexe des transformations de Fourier, reliant la fonction de corrélation et la densité spectrale *).

On peut obtenir les formules (17.4.11) et (17.4.12) directement à partir des formules (17.3.9) et (17.3.10) en introduisant la substitution:

$$\cos \omega\tau = \frac{e^{i\omega\tau} + e^{-i\omega\tau}}{2}$$

en posant de plus $S_x(\omega) = 2S_x^*(\omega)$ et en étendant le domaine d'intégration sur l'intervalle $[-\infty, +\infty]$.

En posant $\tau = 0$ dans la formule (17.4.11) on obtient l'expression de la variance de la fonction aléatoire $\overset{\circ}{X}(t)$:

$$D_x = \int_{-\infty}^{\infty} S_x^*(\omega)\, d\omega. \qquad (17.4.13)$$

*) Rappelons que dans le chapitre 13 nous avons déjà rencontré des transformations de Fourier exprimant la fonction caractéristique et la densité de probabilité l'une en fonction de l'autre.

La formule (17.4.13) donne la variance d'une fonction aléatoire sous la forme d'une somme de variances élémentaires réparties avec une certaine densité sur toute la gamme de fréquences de $-\infty$ à $+\infty$.

En comparant la formule (17.4.13) avec la formule (17.3.2) pour la forme réelle du développement spectral, on voit qu'elles ne diffèrent que par la densité spectrale $S_x^*(\omega)$ qui dans la formule (17.4.13)

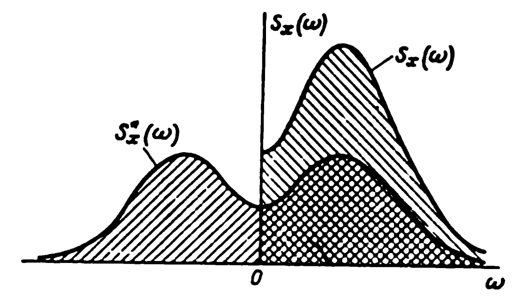


Fig. 17.4.2

est définie non pas de 0 à ∞, mais de $-\infty$ à $+\infty$, en revanche ses ordonnées sont deux fois moindres. Les deux fonctions de la densité spectrale sont représentées sur un graphique (fig. 17.4.2). On voit qu'elles ne diffèrent que par l'échelle suivant l'axe des ordonnées, et par le fait que la fonction $S_x(\omega)$ n'est pas définie pour les fréquences négatives. En pratique on utilise tant l'une que l'autre expression de la densité spectrale.

Parfois on prend pour argument de la densité spectrale non pas la pulsation ω mais la fréquence des oscillations exprimées en hertz:

$$f = \frac{\omega}{2\pi}.$$

Dans ce cas en posant $\omega = 2\pi f$ la formule (17.4.11) devient

$$k_x(\tau) = 2\pi \int_{-\infty}^{\infty} S_x^*(2\pi f)\, e^{2\pi i f \tau}\, df,$$

ou en introduisant la désignation:

$$G_x(f) = 2\pi S_x^*(2\pi f),$$

on a définitivement:

$$k_x(\tau) = \int_{-\infty}^{\infty} G_x(f)\, e^{2\pi i f \tau}\, df. \qquad (17.4.14)$$

La fonction $G_x(f)$ peut également être utilisée comme la densité spectrale de la variance. Elle s'exprime comme suit à l'aide de la

fonction de corrélation:

$$G_x(f) = \int_{-\infty}^{\infty} k_x(\tau)\, e^{-2\pi i f \tau}\, d\tau. \qquad (17.4.15)$$

Les expressions que nous avons données pour la densité de probabilité, ainsi que certaines autres utilisées dans la pratique, ne diffèrent que par l'échelle. Chacune d'elles peut être normée en divisant la densité spectrale correspondante par la variance de la fonction aléatoire.

Exemple 1. La fonction de corrélation de la fonction aléatoire $X(t)$ est donnée par la formule:

$$k_x(\tau) = D e^{-\alpha|\tau|}, \qquad (17.4.16)$$

où $\alpha > 0$ (fig. 17.4.3).

En utilisant la transformation de Fourier sous forme complexe trouver la densité spectrale $S_x^*(\omega)$.

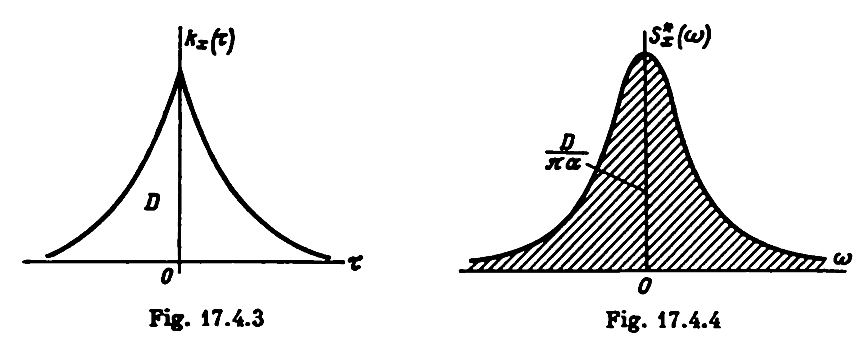


Fig. 17.4.3

Fig. 17.4.4

Solution. La formule (17.4.12) donne:

$$S_x^*(\omega) = \frac{D}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha|\tau|} e^{-i\omega\tau}\, d\tau = \frac{D}{2\pi}\left\{\int_{-\infty}^{0} e^{\alpha\tau} e^{-i\omega\tau}\, d\tau + \int_{0}^{\infty} e^{-\alpha\tau} e^{-i\omega\tau}\, d\tau\right\} =$$

$$= \frac{D}{2\pi}\left\{\int_{0}^{\infty} e^{-(\alpha-i\omega)\tau}\, d\tau + \int_{0}^{\infty} e^{-(\alpha+i\omega)\tau}\, d\tau\right\} =$$

$$= \frac{D}{2\pi}\left[\frac{-1}{\alpha-i\omega} e^{-(\alpha-i\omega)\tau} + \frac{-1}{\alpha+i\omega} e^{-(\alpha+i\omega)\tau}\right]_0^{\infty} =$$

$$= \frac{D}{2\pi}\left[\frac{1}{\alpha-i\omega} + \frac{1}{\alpha+i\omega}\right] = \frac{D\alpha}{\pi(\alpha^2+\omega^2)}.$$

La densité spectrale:

$$S_x^*(\omega) = \frac{D\alpha}{\pi(\alpha^2+\omega^2)}$$

est représentée sur la figure 17.4.4.

Nous allons étudier la fonction de corrélation et la densité spectrale pour différentes valeurs de α.

Pour des α petits la fonction de corrélation décroît plus lentement et l'allure de la fonction aléatoire devient plus douce; les fréquences basses dominent dans le spectre de la fonction aléatoire; la courbe de la densité spectrale s'allonge vers le haut, se rétrécissant simultanément des côtés; à la limite pour $\alpha \to 0$ la fonction aléatoire dégénère en une variable aléatoire à spectre discret se composant d'une seule raie à la pulsation $\omega_0 = 0$.

Au fur et à mesure de l'augmentation de α la fonction de corrélation décroît rapidement, les oscillations de la fonction aléatoire deviennent brusques

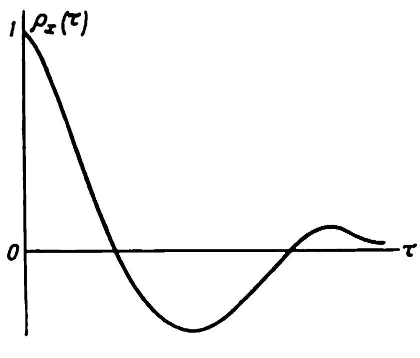


Fig. 17.4.5

et désordonnées; dans le spectre de la fonction aléatoire l'influence des fréquences basses se fait de moins en moins sentir; pour $\alpha \to \infty$ le spectre de la fonction aléatoire tend à être uniforme (celui du « bruit blanc ») sans fréquences prédominantes.

Exemple 2. Soit la fonction de corrélation normée de la fonction aléatoire $\overset{\circ}{X}(t)$ (fig. 17.4.5):

$$\rho_x(\tau) = e^{-\alpha|\tau|} \cos \beta\tau$$

Trouver la densité spectrale normée.

Solution. Ecrivons $\rho_x(\tau)$ sous forme complexe:

$$\rho_x(\tau) = e^{-\alpha|\tau|} \frac{e^{i\beta\tau} + e^{-i\beta\tau}}{2}.$$

La densité spectrale normée $s_x^*(\omega)$ est donnée par la formule (17.4.12); en y substituant $\rho_x\tau$ au lieu de $k_x(\tau)$ on a:

$$s_x^*(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\alpha|\tau|} \frac{e^{i\beta\tau} + e^{-i\beta\tau}}{2} e^{-i\omega\tau} d\tau =$$

$$= \frac{1}{4\pi} \left\{ \int_{\infty}^{0} e^{\alpha\tau} (e^{i\beta\tau} + e^{-i\beta\tau}) e^{-i\omega\tau} d\tau + \int_{0}^{\infty} e^{-\alpha\tau} (e^{i\beta\tau} + e^{-i\beta\tau}) e^{-i\omega\tau} d\tau \right\},$$

d'où après des transformations élémentaires:

$$s_x^*(\omega) = \frac{1}{2\pi} \left\{ \frac{\alpha}{\alpha^2 + (\omega+\beta)^2} + \frac{\alpha}{\alpha^2 + (\omega-\beta)^2} \right\}.$$

L'allure de la courbe de la densité spectrale dépend du rapport des paramètres α et β et est différente suivant que prédomine dans la fonction de corrélation l'amortissement de loi $e^{-\alpha|\tau|}$ ou l'oscillation de loi $\cos \beta\tau$. Il est clair que quand $\alpha < \beta$ le processus est manifestement oscillatoire, pour des $\alpha > \beta$ c'est l'amortissement qui prédomine.

Dans le premier cas la fonction aléatoire est presque périodique de pulsation β, d'amplitude et de phase aléatoires; dans le spectre de la fonction aléatoire

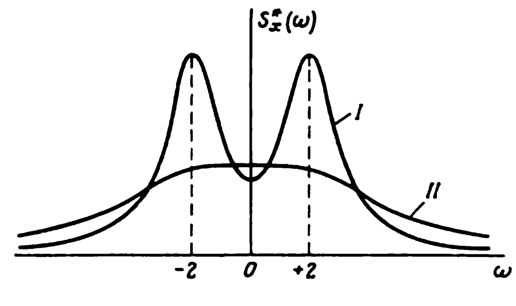


Fig. 17.4.6

prévalent les pulsations voisines de β. Dans le second cas la composition spectrale de la fonction aléatoire est plus uniforme sans influence prononcée de fréquences particulières; à la limite pour $\alpha \to \infty$ le spectre de la fonction aléatoire tend vers celui du bruit blanc.

A titre d'illustration sur la figure 17.4.6 on a représenté les densités de probabilité normées pour les cas suivants: 1) $\beta = 2$, $\alpha = 1$ (courbe *I*); 2) $\beta = 2$, $\alpha = 3$ (courbe *II*). La figure montre que pour $\alpha = 1$ le spectre de la fonction aléatoire accuse des maxima au voisinage des pulsations $\omega = \pm \beta$. Pour $\alpha = 3$ (courbe *II*) la densité spectrale dans une large gamme de fréquences reste presque constante.

17.5. Transformation d'une fonction aléatoire stationnaire par un système linéaire stationnaire

Dans le chapitre 16 nous avons étudié les règles générales des transformations linéaires des fonctions aléatoires données sous forme de développements canoniques. En vertu de ces règles, lors des transformations linéaires des fonctions aléatoires, leurs espérances mathématiques et leurs fonctions de coordonnées sont soumises aux mêmes transformations linéaires. Ainsi, soumettre à une transformation linéaire une fonction aléatoire équivaut à appliquer cette même transformation linéaire à plusieurs fonctions non aléatoires.

Dans le cas des transformations linéaires des fonctions aléatoires stationnaires, le problème se simplifie davantage. Si l'action d'entrée $\overset{\circ}{X}(t)$ et la réponse du système $\overset{\circ}{Y}(t)$ sont stationnaires, la transformation d'une fonction aléatoire peut se réduire à la transformation d'une seule fonction non aléatoire, à savoir de la densité spectrale $S_x(\omega)$.

Pour que la réponse à une action stationnaire le soit également, il faut évidemment que les paramètres du système (par exemple, les résistances, les capacités, les inductances des circuits, etc.) soient constants et non variables. Nous appellerons *système linéaire stationnaire* un système linéaire à paramètres constants. Généralement, le fonctionnement d'un système linéaire stationnaire est décrit par des équations différentielles linéaires à coefficients constants.

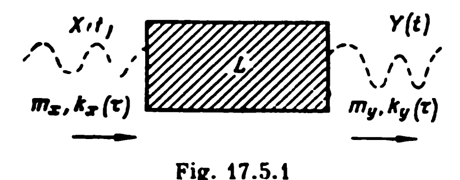


Fig. 17.5.1

Nous allons maintenant étudier la transformation d'une fonction aléatoire stationnaire par un système linéaire stationnaire. Supposons qu'une fonction aléatoire stationnaire $X(t)$ d'espérance mathématique m_x et de fonction de corrélation $k_x(\tau)$ soit appliquée à un système linéaire L. La fonction aléatoire $Y(t)$ est la réponse du système (fig. 17.5.1) dont il y a lieu de déterminer les caractéristiques.

Comme pour la solution du problème posé il nous faudra transformer des fonctions non aléatoires, à savoir l'espérance mathématique et les fonctions de coordonnées, nous déterminons tout d'abord la réponse du système L à une action non aléatoire $x(t)$.

Ecrivons sous forme opérationnelle l'équation différentielle à coefficients constants reliant la réponse $y(t)$ du système à l'action $x(t)$:

$$(a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0)\, y(t) = \\ = (b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0)\, x(t), \qquad (17.5.1)$$

où $p = \frac{d}{dt}$ est l'opérateur de dérivation.

En employant l'écriture condensée de (17.5.1):

$$A_n(p)\; y(t) = B_m(p)\; x(t), \qquad (17.5.2)$$

et en résolvant formellement l'équation (17.5.2) par rapport à $y(t)$ on peut exprimer sous forme explicite l'opérateur du système:

$$y(t) = \frac{B_m(p)}{A_n(p)}\, x(t). \qquad (17.5.3)$$

La solution de l'équation différentielle linéaire (17.5.1) fournit la réponse du système L à l'action $x(t)$. Il est connu de la théorie des équations différentielles que cette solution se compose de deux termes $y_I(t)$ et $y_{II}(t)$. Le terme $y_{II}(t)$ est la solution de l'équation

sans second membre et donne les oscillations *propres* ou *libres* du système. Ce sont les oscillations du système en l'absence d'action à l'entrée, si à l'instant initial le système s'est trouvé hors de l'état d'équilibre. Dans la pratique le plus souvent on a affaire à des systèmes stables; les oscillations propres de ces systèmes s'amortissent dans le temps.

Si l'on considère les intervalles de temps suffisamment éloignés de l'origine du processus, lorsque tous les processus transitoires sont terminés et le système travaille en régime stable, on peut omettre le second terme $y_{II}(t)$ en se limitant au premier terme $y_I(t)$. Ce premier terme traduit les oscillations *forcées* du système dues à l'action donnée $x(t)$.

Dans le cas où l'action $x(t)$ est une fonction analytique assez simple, on arrive souvent à trouver la réponse sous une forme analytique également simple. En particulier, lorsque l'action est une oscillation harmonique de fréquence donnée, la sortie sera également une oscillation harmonique de même fréquence, mais d'amplitude et de phase différentes.

Comme les fonctions de coordonnées du développement spectral d'une fonction aléatoire stationnaire $X(t)$ sont des oscillations harmoniques, il nous faut tout d'abord savoir déterminer la réponse du système à une oscillation harmonique de fréquence donnée ω. La solution de ce problème est particulièrement simple si l'oscillation harmonique s'écrit sous forme complexe.

Supposons qu'une oscillation de la forme:

$$x(t) = e^{i\omega t} \tag{17.5.4}$$

agisse à l'entrée du système.

Cherchons la réponse du système $y(t)$ également sous la forme d'une oscillation harmonique de fréquence ω multipliée par un certain facteur complexe $\Phi(i\omega)$:

$$y(t) = \Phi(i\omega)\, e^{i\omega t}. \tag{17.5.5}$$

Calculons le facteur $\Phi(i\omega)$. A cet effet nous allons substituer la fonction (17.5.4) dans le second membre et la fonction (17.5.5) dans le premier membre de l'équation (17.5.1), il vient:

$$a_n \frac{d^n}{dt^n}[\Phi(i\omega)\, e^{i\omega t}] + a_{n-1}\frac{d^{n-1}}{dt^{n-1}}[\Phi(i\omega)\, e^{i\omega t}] + \ldots$$

$$\ldots + a_1 \frac{d}{dt}[\Phi(i\omega)\, e^{i\omega t}] + a_0[\Phi(i\omega)\, e^{i\omega t}] =$$

$$= b_m \frac{d^m}{dt^m} e^{i\omega t} + b_{m-1}\frac{d^{m-1}}{dt^{m-1}} e^{i\omega t} + \ldots + b_1 \frac{d}{dt} e^{i\omega t} + b_0 e^{i\omega t}. \tag{17.5.6}$$

Comme pour k quelconque on a

$$\frac{d^k}{dt^k} e^{i\omega t} = (i\omega)^k e^{i\omega t}, \quad \frac{d^k}{dt^k} [\Phi(i\omega) e^{i\omega t}] = (i\omega)^k e^{i\omega t} \Phi(i\omega),$$

en divisant les deux membres de l'équation (17.5.6) par $e^{i\omega t}$, on obtient

$$\Phi(i\omega) [a_n (i\omega)^n + a_{n-1} (i\omega)^{n-1} + \ldots + a_1 (i\omega) + a_0] = = b_m (i\omega)^m + b_{m-1} (i\omega)^{m-1} + \ldots + b_1 (i\omega) + b_0. \quad (17.5.7)$$

Le facteur de $\Phi(i\omega)$ n'est donc rien d'autre que le polynôme $A_n(p) = a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0$ où l'opérateur p a été remplacé par $(i\omega)$; d'une manière analogue, le second membre de l'égalité (17.5.7) est l'expression de $B_m(i\omega)$. L'équation (17.5.7) s'écrit alors

$$\Phi(i\omega) A_n(i\omega) = B_m(i\omega),$$

d'où

$$\Phi(i\omega) = \frac{B_m(i\omega)}{A_n(i\omega)}. \quad (17.5.8)$$

La fonction $\Phi(i\omega)$ est appelée *caractéristique fréquentielle* ou réponse en fréquence d'un système linéaire. Pour trouver la caractéristique fréquentielle il suffit de remplacer l'opérateur de dérivation p par $i\omega$ dans l'opérateur du système écrit sous la forme explicite (17.5.3).

Ainsi, *si l'entrée d'un système linéaire à paramètres constants est attaquée par une oscillation harmonique* $e^{i\omega t}$, *la réponse du système est cette même oscillation harmonique multipliée par la caractéristique fréquentielle du système* $\Phi(i\omega)$. Soit l'action d'entrée

$$x(t) = U e^{i\omega t}, \quad (17.5.9)$$

où U est une certaine grandeur ne dépendant pas de t. Le système étant linéaire, la grandeur U peut être mise en facteur devant le signe de l'opérateur, et la réponse du système à l'action (17.5.9) sera

$$y(t) = U\Phi(i\omega) e^{i\omega t}. \quad (17.5.10)$$

Il est évident que cette propriété se conserve même si la grandeur U est aléatoire (il suffit qu'elle ne dépende pas de t).

Nous allons appliquer les méthodes exposées de transformation des oscillations harmoniques par un système linéaire à l'espérance mathématique de la fonction aléatoire $X(t)$ et aux fonctions de coordonnées de son développement spectral.

Ecrivons l'espérance mathématique m_x de la fonction aléatoire stationnaire $X(t)$ comme une oscillation harmonique de pulsation $\omega = 0$ et posons $\omega = 0$ dans la formule (17.5.8)

$$\Phi(0) = \frac{B_m(0)}{A_n(0)} = \frac{b_0}{a_0}, \quad (17.5.11)$$

d'où l'espérance mathématique à la sortie du système est

$$m_y = \frac{b_0}{a_0} m_x. \qquad (17.5.12)$$

Passons maintenant à l'étude de la transformation linéaire de la partie essentiellement aléatoire

$$\overset{\circ}{X}(t) = X(t) - m_x \qquad (17.5.13)$$

de la fonction $X(t)$.

A cet effet écrivons le développement spéctral de la fonction $\overset{\circ}{X}(t)$ sur l'intervalle $(0, T)$

$$\overset{\circ}{X}(t) = \sum_{k=-\infty}^{\infty} U_k e^{i\omega_k t}, \qquad (17.5.14)$$

où U_k sont des variables aléatoires non corrélées, dont les variances forment le spectre de la fonction aléatoire $X(t)$.

Soit un terme de cette somme

$$X_k(t) = U_k e^{i\omega_k t}. \qquad (17.5.15)$$

La réponse correspondante du système sera

$$Y_k(t) = U_k \Phi(i\omega_k) e^{i\omega_k t}. \qquad (17.5.16)$$

En vertu du principe de superposition, la réponse du système à la somme des actions est égale à la somme des réponses à des actions séparées. Par conséquent, la réponse du système à l'action (17.5.14) peut être écrite comme le développement spectral

$$\overset{\circ}{Y}(t) = \sum_{k=-\infty}^{\infty} U_k \Phi(i\omega_k) e^{i\omega_k t},$$

ou, en introduisant la désignation $U_k \Phi(i\omega_k) = W_k$,

$$\overset{\circ}{Y}(t) = \sum_{k=-\infty}^{\infty} W_k e^{i\omega_k t}, \qquad (17.5.17)$$

W_k étant des fonctions aléatoires non corrélées d'espérances mathématiques nulles.

Nous allons maintenant trouver le spectre de ce développement. A cet effet calculons d'abord la variance de la variable aléatoire complexe W_k dans le développement (17.5.17). La variance d'une variable aléatoire complexe étant égale à l'espérance mathématique du carré de son module, on a

$$D[W_k] = M[\,|U_k \Phi(i\omega_k)^2|\,] = M[\,|U_k|^2 |\Phi(i\omega_k)|^2] =$$
$$= |\Phi(i\omega_k)|^2 M[|U_k|^2] = |\Phi(i\omega_k)|^2 D_k. \qquad (17.5.18)$$

Ce qui nous permet de conclure que *lors de la transformation d'une fonction aléatoire stationnaire par un système stationnaire linéaire,*

chacune des ordonnées du spectre se trouve multipliée par le carré du module de la réponse en fréquence du système pour la fréquence correspondante.

Ainsi, lorsqu'une fonction aléatoire stationnaire traverse un système stationnaire linéaire son spectre se trouve modifié : certaines fréquences sont amplifiées, d'autres au contraire affaiblies (filtrées). Le carré du module de la réponse en fréquence montre comment le système réagit à une oscillation de telle ou telle fréquence.

Tout comme précédemment nous allons passer dans la représentation spectrale d'une fonction aléatoire à la limite pour $T \to \infty$ et du spectre discret à la densité spectrale. Il est évident que la densité spectrale à la sortie du système linéaire s'obtient à partir de la densité spectrale à l'entrée par multiplication par $|\Phi(i\omega)|^2$, tout comme étaient obtenues les ordonnées du spectre discret:

$$S_y(\omega) = |\Phi(i\omega)|^2 S_x(\omega). \tag{17.5.19}$$

D'où la règle relativement simple suivante:

Lors de la transformation par un système linéaire stationnaire d'une fonction aléatoire stationnaire sa densité spectrale se trouve multipliée par le carré du module de la réponse en fréquence du système.

Cette règle permet de trouver facilement les caractéristiques d'une fonction aléatoire à la sortie d'un système linéaire, connaissant les caractéristiques d'entrée.

Soit une fonction aléatoire stationnaire $X(t)$ d'espérance mathématique m_x et de fonction de corrélation $k_x(\tau)$ appliquée à l'entrée d'un système linéaire stationnaire dont l'opérateur est donné par (17.5.3). Il y a lieu de trouver l'espérance mathématique m_y et la fonction de corrélation $k_y(\tau)$ de la fonction aléatoire $Y(t)$ à la sortie du système.

Nous allons résoudre le problème dans l'ordre suivant:

1. Déterminons l'espérance mathématique à la sortie:

$$m_y = \frac{b_0}{a_0} m_x. \tag{17.5.20}$$

2. Cherchons la densité spectrale à l'entrée d'après la fonction de corrélation $k_x(\tau)$ (formule (17.4.12)):

$$S_x(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k_x(\tau) e^{-i\omega\tau} d\tau \;^{*}). \tag{17.5.21}$$

3. A l'aide de la formule (17.5.8) nous calculons la réponse en fréquence du système et le carré de son module:

$$|\Phi(i\omega)|^2 = \frac{|B_m(i\omega)|^2}{|A_n(i\omega)|^2}. \tag{17.5.22}$$

---

*) Pour simplifier l'écriture nous omettons le signe * dans la notation de la densité spectrale.

4. En multipliant la densité spectrale à l'entrée par le carré du module de la réponse en fréquence nous trouvons la densité spectrale à la sortie:

$$S_y(\omega) = |\Phi(i\omega)|^2 S_x(\omega). \tag{17.5.23}$$

5. Enfin, d'après la densité spectrale $S_y(\omega)$ nous obtenons la fonction de corrélation $k_y(\tau)$ à la sortie du système:

$$k_y(\tau) = \int_{-\infty}^{\infty} S_y(\omega)\, e^{i\omega\tau}\, d\omega. \tag{17.5.24}$$

Le problème posé se trouve ainsi résolu.

Souvent il y a lieu de connaître non pas la fonction de corrélation $k_y(\tau)$ à la sortie du système, mais seulement la variance D_y:

$$D_y = k_y(0).$$

La formule (17.5.24) conduit pour $\tau = 0$ à une formule bien plus simple:

$$D_y = \int_{-\infty}^{\infty} S_y(\omega)\, d\omega,$$

ou tenant compte du fait que la fonction $S_y(\omega)$ est paire, on a:

$$D_y = 2\int_{0}^{\infty} S_y(\omega)\, d\omega. \tag{17.5.25}$$

Exemple. Le fonctionnement d'un système dynamique linéaire est décrit par l'équation différentielle linéaire du premier ordre:

$$(a_1p + a_0)\, y(t) = (b_1p + b_0)\, x(t), \tag{17.5.26}$$

ou

$$y(t) = \frac{b_1p + b_0}{a_1p + a_0}\, x(t).$$

L'entrée du système est attaquée par une fonction aléatoire stationnaire $X(t)$ d'espérance mathématique m_x et de fonction de corrélation

$$k_x(\tau) = D_x e^{-\alpha|\tau|}, \tag{17.5.27}$$

où α est un coefficient positif (voir exemple 1 du § 17.4). Trouver l'espérance mathématique m_y et la variance D_y à la sortie du système *).

*) Choisissant pour la fonction de corrélation de la fonction aléatoire $X(t)$ l'expression (17.5.27) fréquemment utilisée dans la pratique pour sa simplicité, nous nous rendons compte qu'en toute rigueur la fonction aléatoire $X(t)$ dont la fonction de corrélation est donnée par cette expression n'est pas dérivable, par conséquent elle ne peut satisfaire à des équations différentielles dans le sens général. On peut surmonter cet obstacle si l'on considère l'expression (17.5.27) comme l'approximation de la fonction de corrélation.

S o l u t i o n. La formule (17.5.20) donne:

$$m_y = \frac{b_0}{a_0} m_x.$$

Il est évident que la grandeur m_y ne dépend pas du paramètre α, croît en raison directe de b_0 et décroît en raison inverse de a_0.

La densité spectrale à l'entrée est calculée comme dans l'exemple 1 du § 17.4:

$$S_x(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k_x(\tau) e^{-i\omega\tau} d\tau = \frac{D_x \alpha}{\pi(\alpha^2+\omega^2)}$$

(voir fig. 17.4.4).

A l'aide de la formule (17.5.8) nous trouvons la réponse en fréquence du système:

$$\Phi(i\omega) = \frac{b_1 i\omega + b_0}{a_1 i\omega + a_0}$$

et le carré de son module:

$$|\Phi(i\omega)|^2 = \frac{|b_1 i\omega + b_0|^2}{|a_1 i\omega + a_0|^2} = \frac{b_1^2\omega^2 + b_0^2}{a_1^2\omega^2 + a_0^2}.$$

Puis nous déterminons la densité spectrale à la sortie du système:

$$S_y(\omega) = |\Phi(i\omega)|^2 S_x(\omega) = \frac{D_x}{\pi} \frac{b_1^2\omega^2 + b_0^2}{a_1^2\omega^2 + a_0^2} \frac{\alpha}{\alpha^2+\omega^2}.$$

Ensuite à l'aide de la formule (17.5.25) nous calculons la variance à la sortie:

$$D_y = \frac{2D_x}{\pi} \int_0^{\infty} \frac{b_1^2\omega^2 + b_0^2}{a_1^2\omega^2 + a_0^2} \frac{\alpha}{\alpha^2+\omega^2} d\omega.$$

Pour calculer l'intégrale nous allons écrire l'expression sous l'intégrale à l'aide de fractions simples:

$$\frac{b_1^2\omega^2 + b_0^2}{a_1^2\omega^2 + a_0^2} \frac{\alpha}{\alpha^2+\omega^2} = \frac{A}{a_0^2 + a_1^2\omega^2} + \frac{B}{\alpha^2+\omega^2}$$

et trouver les coefficients:

$$A = \alpha \frac{a_1^2 b_0^2 - a_0^2 b_1^2}{a_1^2\alpha^2 - a_0^2};$$

$$B = \alpha \frac{b_1^2\alpha^2 - b_0^2}{a_1^2\alpha^2 - a_0^2}.$$

Il vient finalement:

$$D_y = D_x \frac{a_1 b_0^2 + a_0 b_1^2 \alpha}{a_0 a_1 (a_1\alpha + a_0)}.$$

En conclusion de ce paragraphe nous mentionnons comment se transforme par un système linéaire une fonction aléatoire stationnaire contenant une variable aléatoire ordinaire additive:

$$X_1(t) = U_0 + X(t), \tag{17.5.28}$$

U_0 étant une variable aléatoire de variance D_0 et $X(t)$ une fonction aléatoire stationnaire.

La réponse du système à l'action $X_1(t)$ est la somme des réponses aux deux termes de (17.5.28). Nous avons déjà calculé la réponse à l'action $X(t)$. Considérant l'action U_0 comme une oscillation harmonique de pulsation $\omega = 0$, il vient en vertu de la formule (17.5.11) :

$$V_0 = \frac{b_0}{a_0} U_0. \tag{17.5.29}$$

Le terme V_0 s'ajoutera simplement à la réponse du système à action $X(t)$.

17.6. Applications de la théorie des processus aléatoires stationnaires aux problèmes d'analyse et de synthèse des systèmes dynamiques

Dans le paragraphe précédent nous avons envisagé la transformation d'une fonction aléatoire stationnaire par un système linéaire et obtenu des méthodes simples permettant de résoudre ce problème. Rappelons que la transformation d'une fonction aléatoire s'est trouvée réduite à une transformation simple (multiplication par le carré du module de la réponse en fréquence) d'une seule fonction, à savoir de la densité spectrale. Cette simplicité extraordinaire de la théorie spectrale des processus aléatoires stationnaires en fait un appareil de choix pour l'étude des systèmes dynamiques linéaires soumis à l'action de perturbations aléatoires (bruits).

Généralement dans bon nombre de problèmes se rencontrant dans la pratique on n'a pas besoin de connaître la fonction de corrélation $k_y(\tau)$ à la sortie du système mais seulement la variance correspondante

$$D_y = k_y(0)$$

caractérisant les erreurs du système dues aux perturbations aléatoires l'attaquant et pouvant dans beaucoup de cas servir de critère de la qualité du fonctionnement du système.

L'étude des systèmes dynamiques par les méthodes de la théorie des fonctions aléatoires a pour but de résoudre deux types de problèmes que l'on peut appeler « directs » et « inverses ».

Le problème direct consiste à étudier le fonctionnement d'un système dynamique linéaire, dont on connaît les paramètres et dont le fonctionnement est décrit par l'équation différentielle linéaire :

$$\begin{aligned}(a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0)\, y(t) &= \\ = (b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0)\, x(t), &\end{aligned} \tag{17.6.1}$$

en présence d'une action aléatoire stationnaire appelée généralement « bruit stationnaire ». A cet effet on étudie tout d'abord le bruit aléatoire, on détermine sa fonction de corrélation et sa composition spectrale. Puis par les méthodes décrites ci-dessus on trouve le spectre

et la variance de la fonction aléatoire à la sortie du système, cette dernière dépendant évidemment tant des caractéristiques de l'action aléatoire d'entrée que des coefficients de l'équation. Ce problème une fois résolu, on peut estimer la précision du fonctionnement du système donné en présence de bruits.

Dans le problème inverse il y a lieu de choisir les coefficients de l'équation (17.6.1) de telle sorte que, pour une composition spectrale du bruit donnée, les bruits à la sortie soient minimaux. Pour des caractéristiques données de la fonction aléatoire (bruit) à l'entrée du système, la variance à la sortie dépend de tout l'ensemble des coefficients de l'équation:

$$D_y = D_y(a_n, a_{n-1}, \ldots, a_1, a_0, b_m, b_{m-1}, \ldots, b_1, b_0).$$

Les coefficients de cette équation dépendent des paramètres constructifs du système dont certains peuvent varier dans les limites assez larges. On peut choisir les valeurs de ces paramètres de façon à rendre minimale la variance D_y.

Il faut noter que dans la plupart des cas on n'arrive pas à satisfaire à cette condition. En effet les expressions que nous avons trouvées pour la fonction de corrélation et la variance à la sortie du système ne sont vraies que pour des instants t suffisamment éloignés de l'origine du processus aléatoire, lorsque tous les processus transitoires dans le système liés à ses oscillations libres ont eu le temps de s'amortir. Dans la pratique, souvent, on se trouve obligé d'utiliser les systèmes dynamiques linéaires (viseurs, calculateurs, systèmes asservis, etc.) durant un laps de temps limité; la vitesse d'amortissement des processus transitoires dans le système dépend de ses paramètres constructifs, c'est-à-dire des coefficients de l'équation (17.6.1). Si l'on choisit ces coefficients de telle sorte que soit minimale la variance à la sortie (pour des instants suffisamment éloignés) ceci entraîne l'apparition à la sortie du système d'autres erreurs dues à ce que les processus transitoires dans le système n'ont pas encore pris fin. Ces erreurs sont appelées erreurs dynamiques.

La durée limitée de fonctionnement des systèmes linéaires et la nécessité de tenir compte des erreurs dynamiques dictent une solution de compromis telle que, d'une part, la variance à la sortie du système soit suffisamment petite et d'autre part, les erreurs dynamiques ne soient pas trop importantes.

Lorsque les paramètres optimaux sont choisis compte tenu tant de la variance que des erreurs systématiques dynamiques, on prend souvent pour critère de précision du fonctionnement du système, le second moment initial α_2 à la sortie du système:

$$\alpha_2 = D_y + m_y^2 \tag{17.6.2}$$

que l'on s'efforce de rendre minimal. Ici D_y est la variance et m_y l'erreur systématique à la sortie du système.

Parfois dans le choix du critère d'estimation du fonctionnement du système on est guidé de son utilisation éventuelle.

Mentionnons encore un problème que pose la construction rationnelle des systèmes dynamiques. Jusqu'à présent nous n'avons parlé que du choix judicieux des coefficients de l'équation (17.6.1), la forme de l'équation étant supposée donnée. Dans la *synthèse des systèmes dynamiques* le problème se pose d'une manière plus générale. En particulier, il s'agit de définir la forme même de l'équation ou plus généralement, l'*opérateur optimal* du système dynamique. Ces problèmes peuvent être résolus à l'aide de la théorie des fonctions aléatoires.

Souvent dans les problèmes pratiques d'analyse et de synthèse des systèmes dynamiques on n'arrive pas à limiter le champ d'études à des processus aléatoires stationnaires et donc à se servir de l'appareil correspondant de la théorie spectrale. Pourtant dans certains cas il suffit simplement de modifier quelque peu cet appareil pour pouvoir également l'appliquer aux processus non stationnaires. On rencontre fréquemment des fonctions aléatoires et des systèmes dynamiques dits « quasi stationnaires » et dont les caractéristiques et paramètres varient dans le temps relativement lentement. Pour l'étude de ces processus aléatoires V. Pougatchev a proposé une méthode différant peu de la méthode spectrale mais pouvant être appliquée dans une gamme plus étendue de conditions *).

17.7. Ergodicité des fonctions aléatoires stationnaires

Soit une certaine fonction aléatoire stationnaire $X(t)$, supposons qu'il ait lieu d'estimer son espérance mathématique m_x et sa fonction de corrélation $k_x(\tau)$. Ci-dessus (voir § 15.4) nous avons exposé les méthodes permettant d'obtenir ces caractéristiques par dépouillement des données expérimentales que sont les réalisations de la fonction aléatoire $X(t)$. Le traitement de ces réalisations permet de trouver les estimations de l'espérance mathématique $\tilde{m}_x(x)$ et de la fonction de corrélation $\tilde{K}_x(t, t')$. Vu le nombre limité d'observations, la fonction $\tilde{m}_x(t)$ ne sera pas rigoureusement constante; il y aura lieu d'en prendre la moyenne qui est une certaine grandeur constante $\tilde{m}_x$ et de la remplacer par cette moyenne; d'une manière analogue, en prenant la moyenne des $\tilde{K}_x(t, t')$ pour différentes valeurs de $\tau = t' - t$ on obtient la fonction de corrélation $\tilde{k}_x(\tau)$.

Cette méthode de traitement est évidemment assez compliquée et de plus doit être réalisée en deux étapes; tout d'abord on trouve

---

*) Voir V. Pougatchev. *Théorie des fonctions aléatoires et ses applications à la théorie de la commande automatique* (en russe). Fizmathguiz, 1962.

les caractéristiques approchées de la fonction aléatoire puis on calcule la moyenne, également approchée, de ces caractéristiques. Il semble alors naturel d'essayer, pour une fonction aléatoire stationnaire, de simplifier ce traitement complexe en supposant que l'espérance mathématique ne dépende pas du temps et la fonction de corrélation de l'origine des temps.

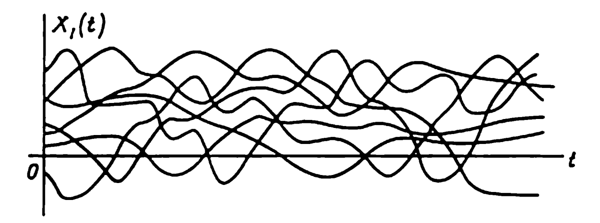


Fig. 17.7.1

Ensuite, il est légitime de s'intéresser s'il est indispensable de disposer de plusieurs réalisations d'une fonction aléatoire stationnaire. Le processus aléatoire étant stationnaire et homogène dans le temps, il est tout naturel de supposer qu'une seule réalisation assez longue suffit pour obtenir les caractéristiques de la fonction aléatoire.

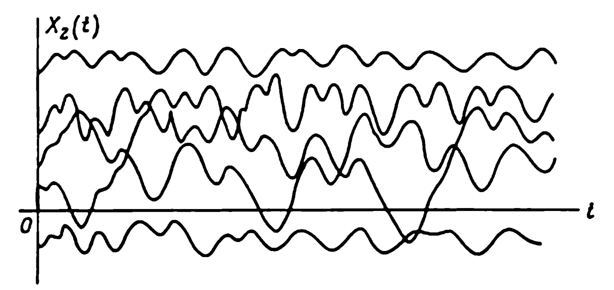


Fig. 17.7.2

Une étude plus détaillée de la question montre que ceci n'est pas vrai pour tous les processus aléatoires; une seule réalisation assez longue n'est pas toujours équivalente à un ensemble de réalisations séparées.

A titre d'exemple nous allons envisager deux fonctions aléatoires stationnaires $X_1(t)$ et $X_2(t)$ (figures 17.7.1 et 17.7.2) données chacune par un ensemble de réalisations.

La fonction aléatoire $X_1(t)$ présente la particularité suivante: toutes ses réalisations ont même valeur moyenne et même amplitude moyenne des oscillations. Choisissons au hasard l'une de ces réalisations et imaginons que l'expérience nous a fourni cette réali-

sation sur un certain intervalle de temps T. Il est évident que si T est suffissamment grand cette réalisation unique donne une information suffisante sur les propriétés de la fonction aléatoire en entier. En particulier, en prenant la moyenne de cette réalisation suivant l'axe des abscisses, donc sur le temps, on espère d'obtenir la moyenne de l'espérance mathématique de la fonction aléatoire; en prenant la moyenne des carrés des écarts de cette moyenne temporelle on obtient la valeur approchée de la variance, etc.

On dit qu'une telle fonction aléatoire est douée de la propriété d'*ergodicité* ou d'*ergodisme*. En vertu de cette propriété chaque réalisation particulière *) de la fonction aléatoire est « un représentant plénipotentiaire » de tout l'ensemble des réalisations éventuelles; une réalisation de durée suffisante équivaut à l'ensemble des réalisations de même durée totale.

Soit maintenant la fonction aléatoire $X_2(t)$. Choisissons au hasard l'une quelconque de ces réalisations que nous supposons durer un temps suffisamment long et calculons sa moyenne temporelle sur tout l'intervalle d'observation. De toute évidence les valeurs moyennes calculées pour différentes réalisations seront différentes et peuvent différer notablement de l'espérance mathématique de la fonction aléatoire calculée comme moyenne d'ensemble. Dans ce cas on dit que la fonction aléatoire n'est pas douée de la propriété d'ergodicité.

Si la fonction aléatoire $X(t)$ est douée de la propriété d'ergodicité, *sa valeur moyenne temporelle (sur un intervalle de temps d'observation suffisant) est approximativement égale à la moyenne d'ensemble d'observations.* Il en sera de même pour $X^2(t)$, $X(t) \times \times X(t+\tau)$, etc. Par conséquent, toutes les caractéristiques de la fonction aléatoire (espérance mathématique, variance, fonction de corrélation) peuvent être calculées d'une manière approchée d'après une seule réalisation suffisamment longue.

Il est donc important de savoir quelles sont les fonctions qui sont douées de la propriété d'ergodicité et quelles sont celles qui ne le sont pas.

Pour élucider cette question nous allons tout d'abord donner un exemple.

Soit une fonction aléatoire $\alpha(t)$ donnant les variations de l'angle d'attaque d'un avion en vol de croisière. Le vol s'effectue dans des conditions météorologiques typiques moyennes. Les variations de l'angle d'attaque sont provoquées par des perturbations aléatoires liées à la turbulence de l'air. La valeur moyenne de l'angle d'attaque dépend de l'altitude de vol H, tout comme l'amplitude des variations

*) En toute rigueur il faudrait dire non pas « chaque réalisation particulière » mais « presque chaque réalisation ». Cette restriction est due à l'éventualité d'apparition de réalisations n'ayant pas cette propriété, mais leur probabilité est nulle.

autour de cette valeur moyenne; en effet, dans les couches basses de l'atmosphère la turbulence est plus forte que dans les couches élevées.

Considérons diverses réalisations de la fonction $\alpha(t)$ pour une altitude donnée H. Soumises à l'action d'un même groupe de facteurs aléatoires, ces réalisations possèdent les mêmes caractéristiques probabilistes, la fonction $\alpha(t)$ est donc ergodique (fig. 17.7.3).

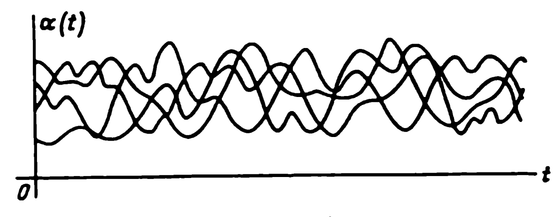


Fig. 17.7.3

Considérons maintenant diverses réalisations de $\alpha(t)$ pour toute une gamme d'altitudes distribuées suivant une certaine loi (par exemple, suivant une loi de densité uniforme). Dans ces conditions, la fonction $\alpha(t)$, encore que stationnaire, ne sera pas ergodique; ses réalisations éventuelles, se produisant avec certaines probabilités, diffèrent par le caractère de variations (fig. 17.7.4).

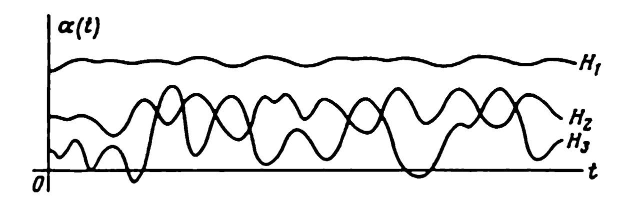


Fig. 17.7.4

Ce processus aléatoire est décomposable en processus aléatoires élémentaires dont chacun a une certaine probabilité de réalisation et les caractéristiques particulières. Cette possibilité de développement et la non-homogénéité interne du processus aléatoire se déroulant de telle ou telle manière avec une certaine probabilité sont la cause physique de la non-ergodicité de ce processus.

En particulier, la non-ergodicité du processus aléatoire peut être liée à la présence dans sa composition d'une variable aléatoire ordinaire (c'est-à-dire à la présence dans le spectre du processus aléatoire en plus d'une partie continue, d'une variance finie à $\omega = 0$).

Soit la fonction aléatoire:

$$Z(t) = X(t) + Y, \tag{17.7.1}$$

où $X(t)$ est une fonction aléatoire stationnaire ergodique de caractéristiques m_x, $k_x(\tau)$ et Y une variable aléatoire de caractéristiques m_y et D_y ; supposons de plus que $X(t)$ et Y ne soient pas corrélées.

Déterminons les caractéristiques de $Z(t)$. En vertu de la loi d'addition des fonctions aléatoires (voir § 15.8) on a:

$$m_z = m_x + m_y, \qquad (17.7.2)$$

$$k_z(\tau) = k_x(\tau) + D_y. \qquad (17.7.3)$$

Les formules (17.7.2) et (17.7.3) montrent que la fonction aléatoire $Z(t)$ est stationnaire. Mais cette fonction est-elle ergodique?

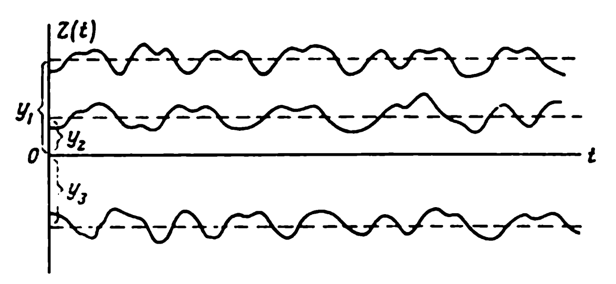


Fig. 17.7.5

L'intuition conduit à répondre « non ». Le caractère de chacune de ses réalisations est différent et leurs moyennes temporelles sont déterminées par la valeur prise par la variable aléatoire Y (fig. 17.7.5).

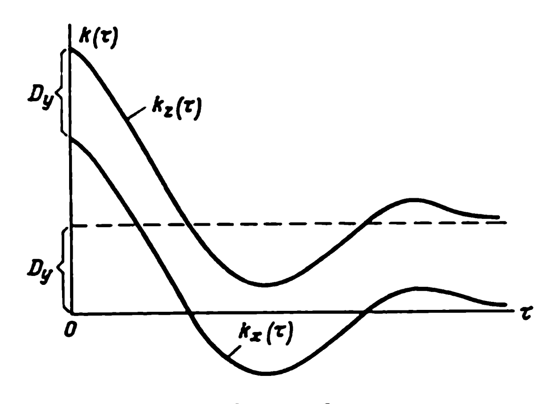


Fig. 17.7.6

La forme de la fonction de corrélation permet de juger directement si un processus aléatoire est ergodique ou non. Soit la fonction de corrélation d'un processus aléatoire non ergodique (17.7.1). Elle diffère de la fonction de corrélation de la fonction aléatoire $X(t)$ par la présence du terme constant D_y (fig. 17.7.6). La fonction de

corrélation $k_x(\tau)$ tend vers zéro pour $t \to \infty$ (la liaison corrélationnelle entre les valeurs de la fonction aléatoire décroît indéfiniment au fur et à mesure de l'augmentation de la distance les séparant), par contre la fonction $k_z(\tau)$ ne tend pas vers zéro pour $t \to \infty$ mais vers la grandeur constante D_y.

Dans la pratique on ne peut pas étudier le processus aléatoire et sa fonction de corrélation durant un temps infini ; le domaine des valeurs de τ que nous pouvons envisager est toujours limité. Si la fonction de corrélation d'un processus aléatoire stationnaire ne décroît pas quand τ augmente, mais à partir d'une certaine valeur de τ reste à peu près constante, ceci signifie, en général, que la fonction aléatoire contient comme composante une variable aléatoire ordinaire et que donc le processus n'est pas ergodique. Au contraire, lorsque la fonction de corrélation tend vers zéro pour $t \to \infty$, ceci témoigne, en général, de l'ergodicité du processus. En tout cas, cette condition est suffisante pour que l'espérance mathématique de la fonction puisse être déterminée comme la moyenne temporelle.

Dans la pratique souvent on établit qu'un processus aléatoire est ergodique non pas en étudiant sa fonction de corrélation pour $t \to \infty$, mais sur la base de considérations physiques, liées à la nature du processus étudié (en particulier, sa « décomposabilité » ou sa « non-décomposabilité » en processus élémentaires de types différents apparaissant avec certaines probabilités).

17.8. Détermination des caractéristiques d'une fonction aléatoire stationnaire ergodique d'après une seule réalisation

Soit une fonction aléatoire stationnaire $X(t)$ ergodique et supposons qu'on dispose d'une seule réalisation de cette fonction aléatoire mais observée sur un intervalle de temps T suffisamment important. Dans le cas d'une fonction aléatoire stationnaire ergodique, une réalisation suffisamment longue est pratiquement équivalente (au sens des propriétés statistiques) à un ensemble de réalisations de même durée totale ; les caractéristiques de la fonction aléatoire peuvent être déterminées non pas comme les moyennes d'ensemble, mais comme des *moyennes temporelles*. En particulier, pour T suffisamment grand, l'espérance mathématique m_x peut approximativement être calculée par la formule

$$m_x \approx \frac{1}{T} \int_0^T x(t)\, dt\ ^{*}). \qquad (17.8.1)$$

---

*) Pour simplifier l'écriture on omet dans les notations des caractéristiques d'une fonction aléatoire le signe ~ signifiant qu'on a affaire non pas à des caractéristiques mais à leurs estimations.

D'une manière analogue, on peut trouver la fonction de corrélation $k_x(\tau)$ pour τ quelconque. En effet, par définition, la fonction de corrélation n'est rien d'autre que *l'espérance mathématique* de la fonction aléatoire $\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)$:

$$k_x(\tau) = M[\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)]. \qquad (17.8.2)$$

De toute évidence, cette espérance mathématique peut également être calculée comme la moyenne temporelle.

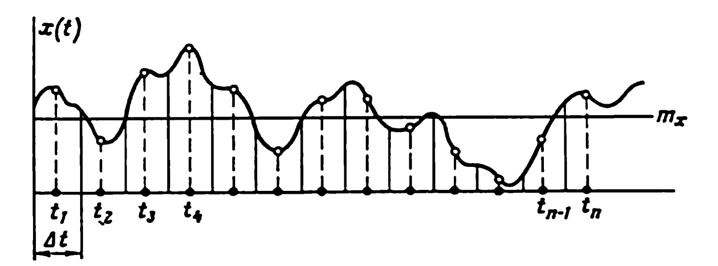


Fig. 17.8.1

Fixons une valeur de τ et calculons comme indiqué la fonction de corrélation $k_x(\tau)$. A cet effet il y a lieu de centrer la réalisation $x(t)$ étudiée, c'est-à-dire d'en retrancher l'espérance mathématique (17.8.1):

$$\overset{\circ}{x}(t) = x(t) - m_x. \qquad (17.8.3)$$

Nous allons calculer pour τ donné l'espérance mathématique de la fonction aléatoire $\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)$ comme la moyenne temporelle. Il nous faudra opérer non pas sur tout l'intervalle de 0 à T, mais sur un segment moindre, car le second facteur $\overset{\circ}{X}(t+\tau)$ n'est défini que pour $t+\tau \leqslant T$.

En calculant la moyenne temporelle par la méthode mentionnée on obtient:

$$k_x(\tau) \approx \frac{1}{T-\tau}\int_0^{T-\tau} \overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)\,dt. \qquad (17.8.4)$$

Ayant calculé l'intégrale (17.8.4) pour différentes valeurs de τ on peut approximativement reproduire point par point la marche de la fonction de corrélation.

Dans la pratique on remplace les intégrales (17.8.1) et (17.8.4) par des sommes finies. Nous allons expliquer ci-dessous la manière de procéder. Divisons l'intervalle d'observation de la fonction aléatoire en n parties égales de longueur Δt et désignons par $t_1, t_2, \ldots, t_n$ les milieux des segments obtenus (fig. 17.8.1). Ecrivons l'intégrale

(17.8.1) comme une somme d'intégrales sur les segments élémentaires Δt mettant en facteur devant le signe de l'intégrale la fonction $x(t)$ égale dans chaque intégrale à la moyenne $x(t_i)$ correspondant au centre sur le segment élémentaire. On a approximativement:

$$m_x = \frac{1}{T}\frac{T}{n}\sum_{i=1}^{n} x(t_i)$$

ou

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x(t_i). \tag{17.8.5}$$

D'une manière analogue, on peut calculer la fonction de corrélation pour des valeurs de τ égales à 0, Δt, $2\Delta t$, . . . Donnons par exemple, à la grandeur τ la valeur

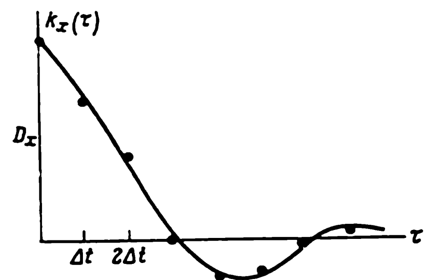


Fig. 17.8.2

$$\tau = m\Delta t = \frac{mT}{n},$$

et calculons l'intégrale (17.8.4) en divisant l'intervalle d'intégration

$$T - \tau = T - \frac{mT}{n} = \frac{n-m}{n} T$$

en $n - m$ segments de même longueur Δt et en mettant dans chacun d'eux en facteur devant l'intégrale la valeur moyenne de la fonction $\overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)$. On obtient:

$$k_x\left(\frac{mT}{n}\right) = \frac{n}{(n-m)T}\frac{T}{n}\sum_{i=1}^{n-m} \overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m}),$$

ou finalement:

$$k_x\left(\frac{mT}{n}\right) = \frac{1}{n-m}\sum_{i=1}^{n-m} \overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m}). \tag{17.8.6}$$

On mène les calculs de la fonction de corrélation à l'aide de la formule (17.8.6) pour $m = 0, 1, 2, \ldots$ successivement jusqu'à des valeurs de m pour lesquelles elle devient pratiquement nulle ou donne lieu à de faibles oscillations non régulières autour de zéro. L'allure générale de la fonction $k_x(\tau)$ se trouve reproduite point par point (fig. 17.8.2).

Pour que l'espérance mathématique m_x et la fonction de corrélation $k_x(\tau)$ puissent être déterminées avec une précision suffisante il faut que le nombre de points n soit suffisamment grand (une cen-

taine et dans certains cas plusieurs centaines). Le choix de la longueur du segment élémentaire Δt se trouve déterminé par le caractère des variations de la fonction aléatoire. Dans le cas des variations relativement lentes, le segment Δt peut être plus grand que si la fonction présente des oscillations brusques et fréquentes. Plus les fréquences des oscillations constituant la fonction aléatoire sont élevées, plus les points repères doivent être serrés. Il est conseillé de partager une période entière de l'harmonique de fréquence la plus élevée en environ 5-10 segments élémentaires Δt.

Souvent le choix des points repères ne dépend pas de l'expérimentateur mais se trouve déterminé par la vitesse de l'appareil enregistreur. Dans ce cas on doit se contenter de traiter les données expérimentales, sans essayer d'introduire de points supplémentaires entre les valeurs observées, car cela n'élèvera pas la précision du résultat mais ne fera que compliquer le traitement.

E x e m p l e. En vcl horizontal d'un avion on a enregistré la surcharge verticale $N(t)$ agissant sur l'avion, ceci pendant 200 s à des intervalles de 2 s. Les résultats sont résumés au tableau 17.8.1. Supposant le processus de variation de la surcharge stationnaire, déterminer approximativement l'espérance mathématique de la surcharge m_N, la variance D_N et la fonction de corrélation

Tableau 17.8.1

t s	Surcharge $N(t)$	t s	Surcharge $N(t)$	t s	Surcharge $N(t)$	t s	Surcharge $N(t)$
0	1,0	50	1,0	100	1,2	150	0,8
2	1,3	52	1,1	102	1,4	152	0,6
4	1,1	54	1,5	104	0,8	154	0,9
6	0,7	56	1,0	106	0,9	156	1,2
8	0,7	58	0,8	108	1,0	158	1,3
10	1,1	60	1,1	110	0,8	160	0,9
12	1,3	62	1,1	112	0,8	162	1,3
14	0,8	64	1,2	114	1,4	164	1,5
16	0,8	66	1,0	116	1,6	166	1,2
18	0,4	68	0,8	118	1.7	168	1,4
20	0,3	70	0,8	120	1,3	170	1,4
22	0,3	72	1,2	122	1,6	172	0,8
24	0,6	74	0,7	124	0,8	174	0,8
26	0,3	76	0,7	126	1,2	176	1,3
28	0,5	78	1,1	128	0,6	178	1,0
30	0,5	80	1,5	130	1,0	180	0,7
32	0,7	82	1,0	132	0,6	182	1,1
34	0,8	84	0,6	134	0,8	184	0,9
36	0,6	86	0,9	136	0,7	186	0,9
38	1,0	88	0,8	138	0,9	188	1,1
40	0,5	90	0,8	140	1,3	190	1,2
42	1,0	92	0,9	142	1,5	192	1,3
44	0,9	94	0,9	144	1,1	194	1,3
46	1,4	96	0,6	146	0,7	196	1,6
48	1,4	98	0,4	148	1,0	198	1,5

normée $\rho_N(\tau)$. Approcher $\rho_N(\tau)$ par une fonction analytique, trouver et représenter graphiquement la densité spectrale du processus aléatoire.

Solution. En vertu de la formule (17.8.5) on a:

$$m_N = \frac{\sum_{i=1}^{100} N(t_i)}{100} \approx 0,98.$$

Puis il y a lieu de procéder au centrage de la fonction aléatoire (tableau 17.8.2).

Tableau 17.8.2

t, s	$\overset{\circ}{N}(t)$	t, s	$\overset{\circ}{N}(t)$	t, s	$\overset{\circ}{N}(t)$	t, s	$\overset{\circ}{N}(t)$
0	0,02	50	0,02	100	0,22	150	−0,18
2	0,32	52	0,12	102	0,42	152	−0,38
4	0,12	54	0,52	104	−0,18	154	−0,08
6	−0,28	56	0,02	106	−0,08	156	0,22
8	−0,28	58	−0,18	108	0,02	158	0,32
10	0,12	60	0,12	110	−0,18	160	−0,08
12	0,32	62	0,12	112	−0,18	162	0,32
14	−0,18	64	0,22	114	0,42	164	0,52
16	−0,18	66	0,02	116	0,62	166	0,22
18	−0,58	68	−0,18	118	0,72	168	0,42
20	−0,68	70	−0,18	120	0,32	170	0,42
22	−0,68	72	0,22	122	0,62	172	−0,18
24	−0,38	74	−0,28	124	−0,18	174	−0,18
26	−0,68	76	−0,28	126	0,22	176	0,32
28	−0,48	78	0,12	128	−0,38	178	0,02
30	−0,48	80	0,52	130	0,02	180	−0,28
32	−0,28	82	0,02	132	−0,38	182	0,12
34	−0,18	84	−0,38	134	−0,18	184	−0,08
36	−0,38	86	−0,08	136	−0,28	186	−0,08
38	0,02	88	−0,18	138	−0,08	188	0,12
40	−0,48	90	−0,18	140	0,32	190	0,22
42	0,02	92	−0,08	142	0,52	192	0,32
44	−0,08	94	−0,08	144	0,12	194	0,32
46	0,42	96	−0,38	146	−0,28	196	0,62
48	0,42	98	−0,58	148	0,02	198	0,52

En élevant au carré toutes les valeurs $\overset{\circ}{N}(t)$ et en divisant la somme des carrés par $n = 100$ on obtient la valeur approchée de la variance de $N(t)$:

$$D_N = \frac{\sum_{i=1}^{100} [\overset{\circ}{N}(t_i)]^2}{100} \approx 0,1045$$

et l'écart quadratique moyen:

$$\sigma_N \approx 0,323.$$

En multipliant entre elles les valeurs $\overset{\circ}{N}(t)$ séparées par l'intervalle $\tau = 2, 4, 6, \ldots$ et en divisant la somme des produits respectivement par $n - 1 = 99$; $n - 2 = 98$; $n - 3 = 97, \ldots$, on obtient les valeurs de la fonction de corrélation $k_N(\tau)$. On procède à la normalisation de la fonction de corrélation en

divisant ses valeurs par $D_N = 0,1045$ ce qui donne le tableau des valeurs de la fonction $\rho_N(\tau)$ (tableau 17.8.3).

Tableau 17.8.3

τ	$\rho_N(\tau)$	$\rho_N^*(\tau) = e^{-\alpha(\tau)}$
0	1,000	1,000
2	0,505	0,598
4	0,276	0,358
6	0,277	0,214
8	0,231	0,128
10	−0,015	0,077
12	0,014	0,046
14	0,071	0,027

Le graphique de la fonction $\rho_N(\tau)$ est donné sur la figure 17.8.3 par des points réunis par une ligne en pointillé. Les irrégularités de la fonction de corrélation peuvent s'expliquer par l'insuffisance des données expérimentales

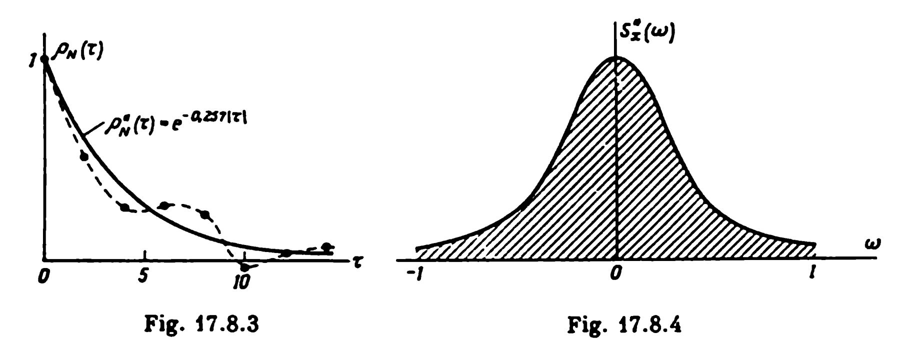


Fig. 17.8.3

Fig. 17.8.4

(durée insuffisante de l'expérience). On continue le calcul de $\rho_N(\tau)$ jusqu'à des valeurs de τ pour lesquelles la relation corrélationnelle ne se manifeste pratiquement plus.

Afin de lisser les oscillations nettement irrégulières de la fonction $\rho_N(\tau)$ trouvée expérimentalement, on approche cette dernière par une fonction du type:

$$\rho_N^*(\tau) = e^{-\alpha|\tau|},$$

où le paramètre α est choisi selon la méthode des moindres carrés (voir § 14.5). Cette méthode donne $\alpha = 0,257$. Calculant les valeurs de la fonction $\rho_N^*(\tau)$ pour $\tau = 0, 2, 4, \ldots$ (dernière colonne du tableau 17.8.3) on construit la courbe approchée (ligne continue sur la figure 17.8.3).

En utilisant pour la fonction de corrélation l'expression approchée (17.8.6) on obtient (voir § 17.4, exemple 1) la densité spectrale normée:

$$S_x^*(\omega) = \frac{\alpha}{\pi(\alpha^2 + \omega^2)} = \frac{0,257}{\pi(0,257^2 + \omega^2)},$$

dont la courbe est donnée par la figure 17.8.4.

CHAPITRE 18

NOTIONS DE BASE DE LA THÉORIE DE L'INFORMATION

18.1. Problèmes de base de la théorie de l'information

La *théorie de l'information* a pour objet d'étude les lois quantitatives liées à l'obtention, la transmission, le traitement et la conservation de l'information. Née dans les années quarante du siècle courant des problèmes pratiques des télécommunications, la théorie de l'information devient à l'heure actuelle l'outil mathématique indispensable pour l'étude des processus de commande les plus divers.

Le caractère aléatoire inhérent aux processus de transmission de l'information implique l'utilisation des méthodes statistiques pour l'étude de ces processus. Pourtant on ne peut pas se limiter aux méthodes classiques de la théorie des probabilités et des notions probabilistes nouvelles doivent être introduites. C'est pourquoi la théorie de l'information n'est pas une simple application des méthodes statistiques mais doit être considérée comme une branche de la théorie des probabilités.

L'obtention, le traitement, la transmission et la conservation de toute sorte d'informations ont lieu dans tout système de commande. Dans le processus de commande il y a toujours échange d'information entre les différentes parties du système. Un exemple simple est la transmission de l'information du dispositif de commande à l'organe d'exécution (transmission des commandes). Un exemple plus compliqué est un circuit fermé de commande où l'information concernant les résultats d'exécution des commandes est transmise au dispositif de commande à l'aide d'une réaction.

Pour être transmise, l'information doit être convenablement « codée », c'est-à-dire traduite en une langue de symboles ou de signaux spéciaux. Les signaux transmettant l'information peuvent être des impulsions électriques, des oscillations lumineuses ou sonores, des déplacements mécaniques, etc.

L'un des problèmes de la théorie de l'information est la recherche des méthodes les plus économiques de codage permettant de transmettre une information donnée à l'aide d'une quantité minimale de symboles. On résout ce problème tant en présence qu'en l'absence de déformations (bruits) dans le canal de transmission.

Dans le second problème typique de la théorie de l'information on dispose d'une source d'information (émetteur) à émission continue et d'un canal de transmission par lequel cette information est transmise à un récepteur. Il y a lieu de trouver le débit de passage du canal de façon à assurer la transmission de toute l'information sans retard ni déformations.

Dans un certain nombre de problèmes de la théorie d'information il s'agit de déterminer la capacité des mémoires destinées à la conservation de l'information, de trouver les méthodes d'introduction et d'extraction de l'information de la mémoire sous la forme convenable pour l'utilisation directe.

Pour pouvoir résoudre ces problèmes il faut avant tout savoir mesurer le volume de l'information transmise et conservée, la capacité de transmission des canaux et leur sensibilité aux bruits (déformations). Les notions de base de la théorie d'information exposées dans le présent chapitre permettent de donner une description quantitative des processus de transmission de l'information et de trouver certaines lois mathématiques régissant ces processus.

18.2. Entropie comme mesure du degré d'incertitude de l'état d'un système physique

Les communications que nous rencontrons dans la théorie de l'information sont toutes un ensemble de données concernant un certain système physique. Par exemple, à l'entrée d'un système de commande automatique d'un atelier industriel peut arriver une communication sur le pourcentage normal ou trop élevé de rebut, sur la composition chimique d'un matériau ou la température d'un four, etc. Toutes ces communications décrivent l'état d'un système physique.

Il est évident que si l'on connaissait l'état d'un système physique toute transmission de l'information serait inutile. La communication n'a de sens que si l'état du système est inconnu d'avance, aléatoire.

Convenons alors que l'information transmise concerne un certain système physique X pouvant se trouver occasionnellement dans un état quelconque, c'est-à-dire un système caractérisé par un certain *degré d'incertitude*. Il est évident que les renseignements que l'on obtiendra sur le système seront d'autant plus importants que l'incertitude du système avant la réception de ces renseignements (« a priori ») était grande. On se pose alors tout naturellement la question suivante : que signifie un degré d'incertitude plus ou moins grand et comment peut-il être mesuré ?

Pour répondre à cette question nous allons tout d'abord comparer entre eux deux systèmes caractérisés chacun par une certaine incertitude.

Soit le premier système une pièce de monnaie qui, jetée en l'air, peut se présenter par l'un des deux côtés : pile ou face. Le second sys-

tème est un dé ayant six états possibles: 1, 2, 3, 4, 5 et 6. Lequel des deux systèmes a l'incertitude plus grande? Il est évident que c'est le second car le nombre de ses états possibles est plus important; notons de plus que ces états sont équiprobables.

On pourrait croire que le degré d'incertitude est déterminé par le nombre d'états possibles du système. Montrons qu'en général ce n'est pas vrai. Considérons, par exemple, un dispositif technique pouvant avoir deux états: bon état de marche et défaillance. Supposons que la probabilité a priori de fonctionnement sans défaillance soit 0,99 et la probabilité de défaillance 0,01. Le degré d'incertitude de ce système est très faible: on peut dire d'une manière presque certaine que le dispositif fonctionnera convenablement. Lorsque l'on jette une pièce de monnaie il y a également deux états possibles mais le degré d'incertitude est bien plus important. On voit ainsi que le degré d'incertitude d'un système physique est déterminé non seulement par le *nombre de ses états possibles*, mais également par la *probabilité des états*.

Passons maintenant au cas général. Soit un système X pouvant prendre un nombre fini d'états $x_1, x_2, \ldots, x_n$ avec les probabilités respectives $p_1, p_2, \ldots, p_n$ où

$$p_i = P(X \sim x_i) \tag{18.2.1}$$

est la probabilité pour le système X de se trouver dans l'état x_i (le symbole $X \sim x_i$ désigne que le système se trouve dans l'état x_i). On a de toute évidence $\sum_{i=1}^{n} p_i = 1$.

Arrangeons ces données sous la forme d'un tableau où dans la ligne supérieure nous marquons les états possibles du système, et dans la ligne inférieure les probabilités correspondantes:

x_i	x_1	x_2	$\ldots$	x_n
p_i	p_1	p_2	$\ldots$	p_n

Ce tableau ressemble à celui de la répartition d'une variable aléatoire discrète X dont les valeurs possibles $x_1, x_2, \ldots, x_n$ ont les probabilités $p_1, p_2, \ldots, p_n$. En effet, il y a beaucoup de commun entre un système physique X à un nombre fini d'états et une variable aléatoire discrète; pour que le premier soit équivalent à la seconde il suffit d'attribuer à chacun des états une certaine valeur numérique (par exemple, le numéro de l'état). Il faut souligner que, pour décrire le degré d'incertitude, peu importe quelles sont les valeurs $x_1, x_2, \ldots, x_n$ qui se trouvent dans la ligne supérieure du tableau, seul importe le *nombre* de valeurs et leur *probabilité*.

Pour la mesure d'incertitude a priori d'un système (ou d'une variable discrète X) on utilise en théorie de l'information une caractéristique spéciale appelée *entropie*. L'entropie est la notion fondamentale de la théorie de l'information.

On appelle entropie d'un système la grandeur

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (18.2.2)$$

qui est la somme des produits des probabilités des différents états du système par les logarithmes de ces probabilités prise avec le signe inverse *).

Comme nous allons le voir l'entropie $H(X)$ jouit d'un certain nombre de propriétés justifiant son utilisation en qualité de caractéristique du degré d'incertitude. Tout d'abord elle s'annule lorsque l'un des états est certain, les autres étant impossibles. Puis, pour un nombre donné d'états elle est maximale lorsque ces états sont équiprobables, et augmente avec le nombre d'états. Enfin, et c'est là peut-être l'essentiel, l'entropie est additive, c'est-à-dire que lorsque plusieurs systèmes indépendants se trouvent réunis en un seul leurs entropies s'ajoutent.

La base a du logarithme dans la formule (18.2.2) peut être quelconque supérieure à 1. Le changement de base équivaut à une simple multiplication de l'entropie par un nombre constant et le choix de la base est le choix d'une certaine unité de mesure de l'entropie. Si l'on choisit pour base le nombre 10, on parle d'unités décimales de l'entropie, si c'est le nombre 2, les unités de mesure sont binaires. En pratique, le plus simple est d'utiliser des logarithmes de base 2 et donc de mesurer l'entropie en unités binaires ; ceci s'accorde bien avec le système binaire de représentation des informations dans les calculatrices électroniques.

Ultérieurement, sauf mention spéciale, par le symbole log on entendra des logarithmes binaires.

Dans l'annexe (table 6) on trouvera les logarithmes binaires des nombres entiers de 1 à 100 **).

Il est facile de voir que lorsque 2 est pris pour base des logarithmes, l'unité de mesure de l'entropie est l'entropie d'un système

---

*) Le signe moins devant la somme a pour but de rendre positive la valeur de l'entropie (les nombres p_i étant inférieurs à l'unité, leurs logarithmes sont négatifs).

**) Les logarithmes des fractions se trouvent par soustraction, par exemple :

$$\log 0{,}13 = \log 13 - \log 100.$$

Pour trouver les logarithmes à trois chiffres significatifs on peut utiliser l'interpolation linéaire.

simple X ayant deux états équiprobables:

x_i	x_1	x_2
p_i	$\frac{1}{2}$	$\frac{1}{2}$

En effet, la formule (18.2.2) donne:

$$H(X) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1.$$

L'unité d'entropie ainsi obtenue est appelée *unité binaire* ou *bit* (de l'anglais « binary digit »—chiffre binaire). C'est l'entropie d'un rang d'un nombre binaire pouvant avec la même probabilité être égal à zéro ou à l'unité.

Nous allons mesurer en unités binaires l'entropie d'un système X ayant n états équiprobables:

x_i	x_1	x_2	...	x_n
p_i	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

$$H(X) = -n\frac{1}{n}\log\frac{1}{n} = -\log 1 + \log n$$

ou

$$H(X) = \log n, \tag{18.2.3}$$

ce qui traduit le fait que *l'entropie d'un système à états équiprobables est égale au logarithme du nombre d'états.*

Par exemple, l'entropie d'un système à huit états est $H(X) = \log 8 = 3$.

Montrons que lorsque l'état du système est connu d'avance, son entropie est nulle. En effet, dans ce cas toutes les probabilités $p_1, p_2, \ldots, p_n$ dans la formule (18.2.2) sont nulles, sauf une, par exemple p_k, qui est égale à l'unité. Le terme $p_k \log p_k$ s'annule car $\log 1 = 0$. Les autres termes sont également nuls car

$$\lim_{p\to 0} p\log p = 0.$$

Montrons que l'entropie du système à un nombre fini d'états est maximale lorsque tous les états sont équiprobables. A cet effet, considérons l'entropie (18.2.2) comme une fonction des probabilités $p_1, p_2, \ldots, p_n$ et trouvons l'extrémum conditionnel de cette fonction pour la condition

$$\sum_{i=1}^{n} p_i = 1. \tag{18.2.4}$$

En utilisant la méthode des multiplicateurs indéterminés de Lagrange cherchons l'extrémum de la fonction

$$F = -\sum_{i=1}^{n} p_i \log p_i + \lambda \sum_{i=1}^{n} p_i. \qquad (18.2.5)$$

En dérivant (18.2.5) par rapport à $p_1, \ldots, p_n$ et en annulant les dérivées on obtient le système suivant d'équations:

$$\log p_i + \log e + \lambda = 0 \quad (i = 1, \ldots, n)$$

ou

$$\log p_i = -\lambda - \log e \quad (i = 1, \ldots, n), \qquad (18.2.6)$$

ce qui montre que l'extrémum (ici le maximum) correspond à des p_i égales entre elles. En vertu de la condition (18.2.4) on a

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n} \qquad (18.2.7)$$

et l'entropie maximale du système

$$H_{\max}(X) = \log n, \qquad (18.2.8)$$

c'est-à-dire que la valeur maximale de l'entropie d'un système à un nombre fini d'états est égale au logarithme du nombre d'états et correspond à des états équiprobables.

Le calcul de l'entropie d'après la formule (18.2.2) peut être quelque peu simplifié par l'introduction de la fonction spéciale

$$\eta(p) = -p \log p, \qquad (18.2.9)$$

où le logarithme est de base 2.

La formule (18.2.2) devient alors:

$$H(X) = \sum_{i=1}^{n} \eta(p_i). \qquad (18.2.10)$$

La fonction $\eta(p)$ est tabulée, ses valeurs pour $0 < p < 1$ sont données de 0,01 en 0,01 dans l'annexe (table 7).

Exemple 1. Un système physique (un appareil) se compose de deux parties I et II. Dans chacune de ces parties un défaut peut survenir indépendamment de l'autre. Après un certain temps de fonctionnement le système peut se trouver dans l'un des quatre états suivants:

1) les deux parties fonctionnent;
2) la partie I fonctionne et la partie II est en défaut;
3) la partie I est en défaut et la partie II fonctionne;
4) les deux parties sont en défaut.

Les probabilités de ces différents états sont respectivement 0,6, 0,25, 0,1 et 0,05. Trouver l'entropie du système.

Solution. Représentons les conditions du problème par un tableau:

x_i	x_1	x_2	x_3	x_4
p_i	0,6	0,25	0,1	0,05

En vertu de la formule (18.2.10) on a:

$$H(X) = \eta(0,6) + \eta(0,25) + \eta(0,1) + \eta(0,05).$$

En utilisant la table 7 de l'annexe on a:

$$H(X) = 0,4422 + 0,5000 + 0,3322 + 0,2161 \approx 1,49 \text{ bit}.$$

Exemple 2. Trouver l'entropie d'un système dont l'état est donné par la loi de répartition d'une variable aléatoire discrète X:

x_i	x_1	x_2	x_3	x_4	x_5
p_i	0,01	0,01	0,01	0,01	0,96

Solution.

$$H(X) = 4\eta(0,01) + \eta(0,96) \approx 0,322 \text{ bit}.$$

Exemple 3. Trouver l'entropie maximale d'un système se composant de trois éléments, pouvant se trouver chacun dans quatre états éventuels.

Solution. Le nombre total d'états possibles du système est égal à $n = 4 \cdot 4 \cdot 4 = 64$. L'entropie maximale du système est égale à $\log 64 = 6$ bits.

Exemple 4. Trouver l'entropie maximale d'une communication de cinq lettres, le nombre de lettres de l'alphabet étant égal à 32.

Solution. Le nombre d'états possibles du système est $n = 32^5$. L'entropie maximale est égale à $5 \log 32 = 25$ bits.

La formule (18.2.2) [ou (18.2.10) qui lui est équivalente] permet de calculer directement l'entropie. Cependant pour les transformations il est plus commode d'écrire l'entropie comme l'espérance mathématique:

$$H(X) = M[-\log P(X)], \tag{18.2.11}$$

où $\log P(X)$ est le logarithme de la probabilité d'un état quelconque (aléatoire) du système, considéré comme une variable aléatoire.

Aux états $x_1, x_2, \ldots, x_n$ du système X correspondent les valeurs suivantes de la variable aléatoire $\log P(X)$:

$$\log p_1, \log p_2, \ldots, \log p_n. \tag{18.2.12}$$

Il est facile de voir que la valeur moyenne (espérance mathématique) de la variable aléatoire $-\log P(X)$ est justement l'entropie du système X. Pour la trouver, il y a lieu de prendre la moyenne pondérée des valeurs (18.2.12), avec des poids égaux aux probabilités correspondantes $p_1, p_2, \ldots, p_n$.

Les formules analogues à (18.2.11) où l'entropie est écrite comme une espérance mathématique permettent de simplifier les opérations impliquant l'entropie en les réduisant à l'emploi familier des théorèmes bien connus sur les espérances mathématiques.

18.3. Entropie d'un système composé. Théorème de composition des entropies

Dans la pratique souvent il y a lieu de déterminer l'entropie d'un système composé obtenu par réunion de deux ou plusieurs systèmes simples.

On entend par *réunion* de deux systèmes X et Y d'états possibles $x_1, \ldots, x_n$; $y_1, \ldots, y_m$ un système composé (X, Y) dont les états (x_i, y_j) sont toutes les combinaisons possibles des états x_i, y_j des systèmes X et Y.

Il est évident que le nombre d'états possibles du système (X, Y) est égal à $n \times m$. Désignons par P_{ij} la probabilité pour le système (X, Y) de se trouver dans l'état (x_i, y_j):

$$P_{ij} = P((X \sim x_i)\ (Y \sim y_j)). \tag{18.3.1}$$

Il est commode de disposer les probabilités P_{ij} sous la forme d'un tableau (matrice):

x_i / y_j	x_1	x_2	...	x_n
y_1	P_{11}	P_{21}	...	P_{n1}
y_2	P_{12}	P_{22}	...	P_{n2}
⋮	⋮	⋮	⋮	⋮
y_m	P_{1m}	P_{2m}	...	P_{nm}

Nous allons calculer maintenant l'entropie d'un système composé. Par définition elle est égale à la somme prise avec le signe inverse des produits des probabilités de tous les états possibles par leurs logarithmes respectifs:

$$H(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log P_{ij} \tag{18.3.2}$$

ou en changeant de désignations:

$$H(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} \eta(P_{ij}). \qquad (18.3.2')$$

L'entropie d'un système composé, tout comme celle d'un système simple, peut s'écrire comme une espérance mathématique, soit:

$$H(X, Y) = M[-\log P(X, Y)], \qquad (18.3.3)$$

où $\log P(X, Y)$, logarithme de la probabilité de l'état du système, est considéré en tant que variable aléatoire (fonction d'état).

Supposons que les systèmes X et Y soient indépendants, c'est-à-dire que leurs états puissent être réalisés indépendamment les uns des autres, et calculons dans cette hypothèse l'entropie du système composé. En vertu du théorème de multiplication des probabilités des événements indépendants on a:

$$P(X, Y) = P(X)\, P(Y),$$

d'où:

$$\log P(X, Y) = \log P(X) + \log P(Y).$$

La substitution dans (18.3.3) donne:

$$H(X, Y) = M[-\log P(X) - \log P(Y)],$$

ou

$$H(X, Y) = H(X) + H(Y), \qquad (18.3.4)$$

c'est-à-dire qu'à *la réunion des systèmes indépendants leurs entropies s'ajoutent*.

La proposition démontrée est le *théorème d'addition des entropies*.

Le théorème d'addition des entropies peut facilement être généralisé à un nombre arbitraire de systèmes indépendants:

$$H(X_1, X_2, \ldots, X_s) = \sum_{k=1}^{s} H(X_k). \qquad (18.3.5)$$

Si les systèmes ne sont pas indépendants le théorème d'addition des entropies n'est plus applicable. Dans ce cas l'entropie d'un système composé est inférieure à l'entropie totale des composantes. Pour trouver l'entropie du système formé d'éléments dépendants il y a lieu d'introduire la notion nouvelle d'*entropie conditionnelle*.

18.4. Entropie conditionnelle. Réunion de systèmes dépendants

Soient deux systèmes X et Y, dans le cas général dépendants l'un de l'autre. Supposons que le système X se trouve dans l'état x_i. Désignons par $P(y_j \mid x_i)$ la probabilité conditionnelle pour le système Y de se trouver dans l'état y_j lorsque le système X se trouve dans l'état

x_i :

$$P(y_j \mid x_i) = P(Y \sim y_j \mid X \sim x_i). \qquad (18.4.1)$$

Déterminons *l'entropie conditionnelle du système Y lorsque le système X se trouve dans l'état x_i*. Désignons-la par $H(Y \mid x_i)$. Par définition :

$$H(Y \mid x_i) = -\sum_{j=1}^{m} P(y_j \mid x_i) \log P(y_j \mid x_i) \qquad (18.4.2)$$

ou

$$H(Y \mid x_i) = \sum_{j=1}^{m} \eta(P(y_j \mid x_i)). \qquad (18.4.2')$$

On peut écrire la formule (18.4.2) sous la forme d'une espérance mathématique, soit :

$$H(Y \mid x_i) = M_{x_i}[-\log P(Y \mid x_i)], \qquad (18.4.3)$$

où M_{x_i} est l'espérance mathématique conditionnelle de la grandeur entre parenthèses, pour la condition $X \sim x_i$.

L'entropie conditionnelle dépend de l'état x_i du système X. Calculons l'entropie *moyenne*, ou *totale*, du système Y compte tenu du fait que le système peut prendre des états différents. A cet effet il y a lieu de multiplier chacune des entropies conditionnelles (18.4.2) par la probabilité de l'état correspondant p_i et d'ajouter tous ces produits. Désignons par $H(Y \mid X)$ l'entropie conditionnelle totale, on a :

$$H(Y \mid X) = \sum_{i=1}^{n} p_i H(Y, x_i) \qquad (18.4.4)$$

ou, compte tenu de la formule (18.4.2) :

$$H(Y \mid X) = -\sum_{i=1}^{n} p_i \sum_{j=1}^{m} P(y_j \mid x_i) \log P(y_j \mid x_i).$$

En portant p_i sous le second signe somme, on a :

$$H(Y \mid X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_i P(y_j \mid x_i) \log P(y_j \mid x_i) \qquad (18.4.5)$$

ou

$$H(Y \mid X) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \eta(P(y_j \mid x_i)). \qquad (18.4.5')$$

En vertu du théorème de multiplication des probabilités on a :

$$p_i P(y_j \mid x_i) = P_{ij},$$

par conséquent :

$$H(Y \mid X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log P(y_j \mid x_i). \qquad (18.4.6)$$

On peut écrire l'expression (18.4.6) comme une espérance mathématique, soit :

$$H(Y \mid X) = M[-\log P(Y \mid X)]. \tag{18.4.7}$$

La grandeur $H(Y \mid X)$ caractérise le degré d'incertitude du système Y connaissant l'état du système X. Nous l'appellerons *entropie totale du système Y conditionnellement à X*.

Exemple. Soient deux systèmes X et Y réunis en un seul (X, Y) dont les probabilités des états sont données par le tableau :

y_j \ x_i	x_1	x_2	x_3	r_j
y_1	0,1	0,2	0	0,3
y_2	0	0,3	0	0,3
y_3	0	0,2	0,2	0,4
p_i	0,1	0,7	0,2	

Trouver les entropies conditionnelles totales $H(Y \mid X)$ et $H(X \mid Y)$.

Solution. En ajoutant les probabilités P_{ij} colonne par colonne on obtient les probabilités $p_i = P(X \sim x_i)$:

$$p_1 = 0{,}1; \quad p_2 = 0{,}7; \quad p_3 = 0{,}2.$$

Ecrivons ces probabilités dans la ligne supplémentaire du tableau. D'une manière analogue, en ajoutant les P_{ij} ligne par ligne on trouve les $r_j = P(Y \sim y_j)$:

$$r_1 = 0{,}3; \quad r_2 = 0{,}3; \quad r_3 = 0{,}4,$$

que l'on porte dans la colonne droite du tableau ci-dessus. En divisant les P_{ij} par p_i on obtient un tableau pour les probabilités conditionnelles $P(y_j \mid x_i)$:

y_j \ x_i	x_1	x_2	x_3
y_1	1	$\frac{0,2}{0,7}$	0
y_2	0	$\frac{0,3}{0,7}$	0
y_3	0	$\frac{0,2}{0,7}$	1

La formule (18.4.5') permet de trouver $H(Y \mid X)$. Les entropies conditionnelles étant nulles pour $X \sim x_1$ et $X \sim x_3$ on a :

$$H(Y \mid X) = 0{,}7\left[\eta\left(\frac{0{,}2}{0{,}7}\right) + \eta\left(\frac{0{,}3}{0{,}7}\right) + \eta\left(\frac{0{,}2}{0{,}7}\right)\right].$$

En utilisant la table 7 de l'annexe on trouve:

$$H(Y \mid X) \approx 1{,}09 \text{ bit.}$$

D'une manière analogue on peut trouver $H(X \mid Y)$. En intervertissant X et Y dans la formule (18.4.5') on a:

$$H(X \mid Y) = \sum_{j=1}^{m} r_j \eta (P(x_i \mid y_j)).$$

Composons le tableau des probabilités conditionnelles $P(x_i \mid y_j)$. En divisant les P_{ij} par r_j on obtient:

y_j \ x_i	x_1	x_2	x_3
y_1	$\frac{0,1}{0,3}$	$\frac{0,2}{0,3}$	0
y_2	0	1	0
y_3	0	$\frac{0,2}{0,4}$	$\frac{0,2}{0,4}$

D'où

$$H(X \mid Y) = 0{,}3\left[\eta\left(\frac{0{,}1}{0{,}3}\right) + \eta\left(\frac{0{,}2}{0{,}3}\right)\right] + 0{,}4\left[\eta\left(\frac{0{,}2}{0{,}4}\right) + \eta\left(\frac{0{,}2}{0{,}4}\right)\right] \approx 0{,}68 \text{ bit.}$$

En utilisant la notion d'entropie conditionnelle on peut exprimer l'entropie d'un système composé en fonction des entropies de ses composantes.

Nous allons démontrer le théorème suivant:

L'entropie d'un système, composé de systèmes X et Y, est égale à l'entropie d'une de ses composantes plus l'entropie conditionnelle de l'autre par rapport à la première:

$$H(X, Y) = H(X) + H(Y \mid X). \qquad (18.4.8)$$

Nous allons écrire $H(X, Y)$ comme l'espérance mathématique (18.3.3):

$$H(X, Y) = M[-\log P(X, Y)].$$

En vertu du théorème de multiplication des probabilités on a:

$$P(X, Y) = P(X)\, P(Y \mid X),$$

par conséquent:

$$\log P(X, Y) = \log P(X) + \log P(Y \mid X),$$

d'où:

$$H(X, Y) = M[-\log P(X)] + M[-\log P(Y \mid X)]$$

ou en vertu des formules (18.2.11), (18.3.3) on a:

$$H(X, Y) = H(X) + H(Y \mid X),$$

ce qu'il fallait démontrer.

En particulier, lorsque les systèmes X et Y sont indépendants, $H(Y \mid X) = H(Y)$ et l'on obtient le théorème d'addition des entropies démontré dans le paragraphe précédent, à savoir:

$$H(X, Y) = H(X) + H(Y).$$

Dans le cas général:

$$H(X, Y) \leqslant H(X) + H(Y). \qquad (18.4.9)$$

La relation (18.4.9) exprime le fait que l'entropie totale conditionnelle $H(Y \mid X)$ ne peut être supérieure à l'entropie inconditionnelle:

$$H(Y \mid X) \leqslant H(Y). \qquad (18.4.10)$$

Nous démontrerons l'inégalité (18.4.10) au § 18.6, bien qu'intuitivement elle soit évidente: le degré d'incertitude d'un système ne peut pas augmenter du fait que l'on connaît l'état d'un autre système.

En vertu de la relation (18.4.9) l'entropie d'un système composé est maximale dans le cas limite, lorsque ses composantes sont indépendantes.

Considérons maintenant l'autre cas limite, lorsque l'état d'un système (par exemple X) détermine entièrement l'état de l'autre système Y. Dans ce cas $H(Y \mid X) = 0$ et la formule (18.4.7) donne:

$$H(X, Y) = H(X).$$

Si l'état de l'un des systèmes X, Y détermine d'une manière univoque l'état de l'autre et inversement (le cas des systèmes équivalents), on a:

$$H(X, Y) = H(X) = H(Y).$$

Le théorème de l'entropie d'un système composé peut être facilement généralisé à un nombre quelconque de systèmes:

$$H(X_1, X_2, \ldots, X_s) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \\ + \ldots + H(X_s \mid X_1, X_2, \ldots, X_{s-1}), \qquad (18.4.11)$$

où l'entropie de chaque système est calculée en supposant que l'état de tous les systèmes précédents soit connu.

18.5. Entropie et information

Dans les paragraphes précédents nous avons défini l'entropie comme la mesure d'incertitude de l'état d'un système physique. Il est évident que si l'on apprend certaine information sur le système

cette incertitude diminue. Plus on a de données, plus ces dernières sont complètes, plus l'information dont on dispose sur le système est grande et moindre est l'incertitude de son état. Il semble donc tout naturel de mesurer la quantité d'information par la diminution de l'entropie du système dont les données en question précisent l'état.

Soit un certain système X; nous allons estimer l'information accumulée au fur et à mesure que l'état du système X devient connu. Avant que l'on ait ces données, l'entropie (a priori) du système était $H(X)$; après l'obtention des données l'état du système s'est entièrement déterminé, c'est-à-dire que l'entropie est devenue égale à zéro. Désignons par I_X l'information obtenue avec la détermination de l'état du système X, évidemment, elle est égale à la diminution de l'entropie:

$$I_X = H(X) - 0$$

ou

$$I_X = H(X), \tag{18.5.1}$$

c'est-à-dire que *la quantité d'information acquise lorsque l'état d'un système physique devient entièrement connu est égale à l'entropie de ce système.*

Écrivons la formule (18.5.1) comme suit:

$$I_X = -\sum_{i=1}^{n} p_i \log p_i, \tag{18.5.2}$$

où $p_i = P(X \sim x_i)$.

La formule (18.5.2) signifie que l'information I_X est la valeur moyenne du logarithme de la probabilité d'état prise sur tous les états du système et avec le signe inverse.

Remarquons que pour obtenir I_X il y a lieu de multiplier $\log p_i$ (logarithme de la probabilité de l'état i) pris avec le signe moins, par la probabilité de cet état puis d'ajouter tous ces produits. Il est tout naturel de considérer chaque terme $-\log p_i$ comme l'information fournie par une communication particulière selon laquelle le système X se trouve dans l'état x_i. Désignons cette information par I_{x_i} :

$$I_{x_i} = -\log p_i. \tag{18.5.3}$$

L'information I_X sera alors l'information *moyenne* (ou *totale*) obtenue de toutes les communications, compte tenu de leurs probabilités. La formule (18.5.2) peut s'écrire comme une espérance mathématique, soit

$$I_X = M[-\log P(X)], \tag{18.5.4}$$

où la lettre X désigne un état quelconque (aléatoire) du système X.

Comme tous les nombres p_i ne sont pas supérieurs à l'unité, ni l'information partielle I_{x_i} ni l'information totale I_X ne peuvent être négatives.

Si tous les états possibles du système sont a priori équiprobables $\left(p_1 = p_2 = \ldots = p_n = \frac{1}{n}\right)$, il est évident que l'information I_{x_i} de chacune des communications particulières

$$I_{x_i} = -\log p = \log n$$

est égale à l'information moyenne (totale):

$$I_X = -n\frac{1}{n}\log\frac{1}{n} = \log n.$$

Dans le cas où les états du système ont des probabilités différentes les informations des diverses communications sont différentes: la plus grande information est contenue dans des communications sur les événements qui, a priori, étaient les moins probables. Par exemple, la communication que le 31 décembre à Moscou il a neigé donne moins d'information qu'une communication identique de par son contenu mais relative au 31 juillet.

Exemple 1. Une pièce du jeu d'échecs se trouve sur une case d'un échiquier. A priori toutes les positions des pièces sont équiprobables. Déterminer l'information contenue dans la communication indiquant la case occupée par la pièce.

Solution. L'entropie du système X à n états équiprobables est égale à $\log n$; ici:

$$I_X = H(X) = \log 64 = 6 \text{ bits},$$

c'est-à-dire que la communication contient 6 bits d'information. Comme tous les états du système sont équiprobables, toute communication concrète du type « la pièce se trouve sur la case e2 » porte la même information.

Exemple 2. Pour les conditions de l'exemple 1, trouver l'information de la communication que la pièce occupe un des coins de l'échiquier.

Solution. La probabilité a priori de l'état dont il est question est:

$$p = \frac{4}{64} = \frac{1}{16}.$$

L'information partielle est égale à:

$$I = -\log\frac{1}{16} = 4 \text{ bits}.$$

Exemple 3. Trouver l'information particulière contenue dans la communication d'une personne A rencontrée pour la première fois: « Aujourd'hui c'est mon anniversaire ».

Solution. A priori, tous les jours de l'année ont la même probabilité d'être les jours d'anniversaire de la personne A. La probabilité de la communication reçue est $p = \frac{1}{365}$. L'information particulière de la communication en question est:

$$I = -\log\frac{1}{365} \approx 8{,}51 \text{ bits}.$$

Exemple 4. Dans les conditions de l'exemple 3 trouver l'information totale de la communication apprenant qu'aujourd'hui est l'anniversaire ou non de la personne A rencontrée pour la première fois.

S o l u t i o n. Le système dont on recherche l'état a deux états possibles: x_1, c'est le jour d'anniversaire et x_2, ce n'est pas le jour d'anniversaire. Les probabilités respectives de ces états sont $p_1 = \frac{1}{365}$; $p_2 = \frac{364}{365}$.

L'information totale est égale à:

$$I_X = H(X) = \eta\left(\frac{1}{365}\right) + \eta\left(\frac{364}{365}\right) \approx 0{,}063 \text{ bit.}$$

E x e m p l e 5. On tire sur une cible n coups indépendants; la probabilité d'atteindre le but avec un coup est égale à p. Les résultats du tir ne peuvent être observés directement. Après le k-ième coup ($1 \leqslant k \leqslant n$) on examine la cible pour voir si elle a été atteinte ou non; si oui on ne tire plus dessus. Trouver k de la condition que la quantité d'information fournie par l'examen de la cible soit maximale.

S o l u t i o n. Soit un système physique X, en l'occurrence la cible, après le k-ième coup. Les états possibles du système X seront:

x_1, la cible a été atteinte;

x_2, la cible n'a pas été atteinte.

Les probabilités des états x_1 et x_2 sont données par le tableau:

x_i	x_1	x_2
p_i	$1-(1-p)^k$	$(1-p)^k$

Il est évident que l'information obtenue lors de l'étude du système X est maximale lorsque les deux états x_1 et x_2 sont équiprobables:

$$1 - (1 - p)^k = (1 - p)^k.$$

D'où

$$2(1 - p)^k = 1; \quad (1 - p)^k = 1/2$$

et

$$k = \frac{-1}{\log(1-p)},$$

en entendant par log le logarithme de base 2. Il est clair que k peut ne pas être un entier, il y a alors lieu de l'arrondir au nombre entier le plus proche. Par exemple, pour $p = 0{,}2$ on a:

$$k = \frac{1}{0{,}3219} \approx 3.$$

Si on obtient $k > n$, ceci signifie que l'information maximale sur l'état du système est fournie par l'examen de la cible après le dernier coup ($k = n$).

Si l'information est mesurée en bits on peut lui donner une interprétation assez simple: c'est le nombre de réponses « oui » et « non » fournissant la même information. En effet, soit un système à deux états:

x_i	x_1	x_2
p_i	p_1	p_2

Pour trouver l'état de ce système il suffit de poser une seule question, par exemple de demander si le système se trouve dans l'état x_1. La réponse « oui » ou « non » à cette question fournit une certaine information, atteignant la valeur maximale 1 lorsque les deux états a priori sont équiprobables $p_1 = p_2 = 1/2$. Ainsi l'information maximale donnée par la réponse « oui » ou « non » est égale à un bit.

Si l'information obtenue d'une communication est égale à n bits, elle est équivalente à l'information donnée par n réponses « oui » ou « non » à des questions posées de telle sorte que les « oui » et « non » soient équiprobables.

Dans certains cas simples pour déterminer le contenu d'une information on arrive à poser des questions de telle sorte que les réponses « oui » et « non » à ces questions soient équiprobables. Dans ces cas l'information reçue se mesure par le nombre de questions posées.

S'il n'est pas possible de poser les questions de cette façon, on peut affirmer seulement que le nombre minimal de questions indispensables pour élucider le contenu de la communication donnée *n'est pas inférieur* à l'information contenue dans la communication. Pour que le nombre de questions soit minimal il y a lieu de les formuler de telle sorte que les probabilités des réponses « oui » ou « non » soient les plus voisines possibles de 1/2.

E x e m p l e 6. Quelqu'un a pensé à un nombre X tel que

$$1 \leqslant X \leqslant 8.$$

Il y a lieu de trouver ce nombre en posant un nombre minimal de questions auxquelles il faudra répondre « oui » ou « non ».

S o l u t i o n. Déterminons l'information contenue dans la communication ci-dessus. A priori toutes les valeurs de X, de 1 à 8, sont équiprobables, donc $p_1 = p_2 = \ldots = p_8 = 1/8$ et la formule (18.5.2) donne:

$$I_X = \log 8 = 3.$$

Le nombre minimal de questions qu'il faudra poser pour trouver le nombre retenu est ainsi non inférieur à trois.

Dans le cas présent trois questions sont effectivement suffisantes, si on les formule de telle sorte que les probabilités de réponses « oui » ou « non » soient égales.

Supposons que c'est le nombre « cinq ». Dans l'ignorance du nombre retenu on pose les questions suivantes:

Q u e s t i o n 1. Le nombre X est-il inférieur à cinq?

R é p o n s e. Non.

(Conclusion: X est l'un des nombres 5, 6, 7, 8).

Q u e s t i o n 2. Le nombre X est-il inférieur à sept?

R é p o n s e. Oui.

(Conclusion: X est l'un des nombres 5, 6).

Q u e s t i o n 3. Le nombre X est-il inférieur à six?

R é p o n s e. Oui.

(Conclusion: X est égal à cinq).

Il est facile de voir que trois questions de ce genre (ou des questions analogues) peuvent permettre de trouver un nombre quelconque entre 1 et 8 *).

Ainsi nous avons appris à mesurer l'information concernant un système X contenue tant dans les communications isolées sur son état que dans le fait même de poser une question relative à cet état. On supposait toujours qu'on observait directement le système X. En fait il n'en ait pas toujours ainsi : il peut se faire que le système X ne soit pas accessible à l'observation ; il y a lieu alors de trouver l'état non pas du système X, mais d'un certain autre système Y lié au premier. Par exemple, au lieu de suivre directement des cibles aériennes, au poste de commande on observe l'écran montrant la situation dans l'air ; sur cet écran les cibles sont représentées par des signes conventionnels. De même, faute de pouvoir suivre directement un navire cosmique on analyse les signaux transmis par son appareillage. Encore, au lieu du texte X d'un télégramme envoyé le destinataire lit le texte Y reçu, celui-ci ne coïncidant pas toujours avec X.

Les écarts entre le système X qui nous intéresse et le système observé Y peuvent dans le cas général être de deux types :

1) Les écarts dus à ce que certains états du système X n'ont pas d'analogues dans le système Y plus pauvre en détails que le système X.

2) Les écarts dus aux erreurs : erreurs de mesure des paramètres du système X et erreurs de transmission des communications.

A titre d'exemple d'écarts du premier type on peut citer les erreurs d'arrondi des données numériques ou les erreurs dues à la description grossière des propriétés du système X par le système Y. Un exemple d'écarts du second type est fourni par les déformations des signaux à cause des bruits dans les canaux de transmission, à cause des défectuosités d'appareillage, d'erreurs d'étourderie des opérateurs participant à la transmission, etc.

Lorsque le système nous intéressant X et le système observé Y sont différents, il y a lieu de connaître *la quantité d'information sur le système X donné par l'observation de Y.*

Il est tout naturel de définir cette information comme une *diminution de l'entropie du système* X avec l'obtention des renseignements concernant l'état du système Y :

$$I_{Y\to X}=H(X)-H(X\mid Y). \qquad (18.5.5)$$

En effet avant la réception des données concernant le système Y l'entropie du système X était $H(X)$; après la réception de ces données l'entropie restante est devenue $H(X\mid Y)$; l'entropie détruite par les données est l'information $I_{Y\to X}$.

*) On recommande au lecteur de poser six questions indispensables pour déterminer la position de la pièce sur un échiquier (voir exemple 1) et de trouver le nombre de questions nécessaires pour déterminer la carte retenue dans un jeu de 36 cartes.

Nous appellerons la quantité (18.5.5) *information totale* (ou moyenne) *sur le système X contenue dans le système Y.*

Nous allons montrer que

$$I_{Y\to X} = I_{X\to Y},$$

c'est-à-dire que chacun des deux systèmes contient sur l'autre la même information totale.

Pour la démonstration nous allons écrire l'entropie du système (X, Y) conformément au théorème (18.4.8) par deux formules équivalentes :

$$H(X, Y) = H(X) + H(Y \mid X),$$

$$H(X, Y) = H(Y) + H(X \mid Y),$$

d'où :

$$H(X) + H(Y \mid X) = H(Y) + H(X \mid Y),$$

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X).$$

ou

$$I_{Y\to X} = I_{X\to Y}, \tag{18.5.6}$$

ce qu'il fallait démontrer.

Introduisons la désignation :

$$I_{Y\leftrightarrow X} = I_{Y\to X} = I_{X\to Y} \tag{18.5.7}$$

et appelons l'information $I_{Y\leftrightarrow X}$ l'*information mutuelle totale* contenue dans les systèmes X et Y.

Voyons maintenant ce que devient l'information mutuelle totale dans les cas limites où les systèmes sont nettement indépendants ou étroitement liés. Si X et Y sont indépendants on a :

$$H(Y \mid X) = H(Y)$$

et

$$I_{Y\leftrightarrow X} = 0, \tag{18.5.8}$$

c'est-à-dire que l'information totale mutuelle contenue dans les systèmes indépendants est nulle. Ce résultat est tout naturel car on ne peut recueillir de renseignements sur un système en observant un autre absolument indépendant du premier.

Envisageons maintenant un autre cas limite lorsque l'état du système X détermine complètement l'état du système Y et inversement (systèmes équivalents). On a alors $H(X) = H(Y)$:

$$H(X \mid Y) = H(Y \mid X) = 0$$

et

$$I_{Y\leftrightarrow X} = I_X = I_Y = H(X) = H(Y), \tag{18.5.9}$$

donc on obtient le cas déjà envisagé [formule (18.5.2)] d'observation directe du système X (ou, ce qui est la même chose, du système équivalent Y).

Considérons le cas où entre X et Y il y a une relation bien déterminée unilatérale : l'état d'un système détermine entièrement celui de l'autre mais non inversement. Nous convenons d'appeler « système dépendant » celui des deux dont l'état est entièrement déterminé par l'état de l'autre. D'après l'état d'un système dépendant il est impossible de déterminer d'une manière univoque l'état de l'autre. Par exemple, si le système X est le texte complet d'une communication littérale, et Y son texte abrégé par omission de toutes les voyelles, en recevant « clt » on ne peut dire avec certitude, si c'est « calot », ou « colite », ou « culte » ou même « acolyte ».

Il est évident que l'entropie d'un système dépendant est inférieure à l'entropie du système auquel il est lié.

Nous allons déterminer l'information mutuelle contenue dans les systèmes dont l'un dépend de l'autre mais non inversement.

Supposons que des deux systèmes X et Y c'est le système X qui soit dépendant. On a alors $H(X \mid Y) = 0$ et

$$I_{Y \leftrightarrow X} = H(X), \tag{18.5.10}$$

ce qui traduit le fait que *l'information mutuelle totale contenue dans les systèmes dont l'un est un système dépendant est égale à l'entropie de ce dernier.*

Nous allons trouver l'expression de l'information $I_{Y \leftrightarrow X}$ non pas en fonction de l'entropie conditionnelle mais directement en fonction de l'entropie $H(X, Y)$ du système composé et des entropies $H(X)$ et $H(Y)$ de ses composantes.

En utilisant le théorème de l'entropie du système composé on obtient :

$$H(X \mid Y) = H(X, Y) - H(Y). \tag{18.5.11}$$

Le report de cette expression dans la formule (18.5.5) donne :

$$I_{Y \leftrightarrow X} = H(X) + H(Y) - H(X, Y), \tag{18.5.12}$$

donc *l'information totale mutuelle contenue dans deux systèmes est égale à la somme des entropies des composantes moins l'entropie du système composé.*

Les relations obtenues permettent facilement de trouver l'expression générale de l'information mutuelle totale sous la forme d'une espérance mathématique. En substituant dans (18.5.12) les expressions

$$H(X) = M[-\log P(X)], \quad H(Y) = M[-\log P(Y)],$$
$$H(X, Y) = M[-\log P(X, Y)],$$

on obtient:

$$I_{Y\leftrightarrow X} = M\left[-\log P(X) - \log P(Y) + \log P(X, Y)\right]$$

ou

$$I_{Y\leftrightarrow X} = M\left[\log \frac{P(X, Y)}{P(X)\,P(Y)}\right] \qquad (18.5.13)$$

Pour le calcul direct de l'information totale mutuelle on écrit la formule (18.5.13) comme suit:

$$I_{Y\leftrightarrow X} = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log \frac{P_{ij}}{p_i r_j}, \qquad (18.5.14)$$

où

$$P_{ij} = P((X \sim x_i)(Y \sim y_j));$$
$$p_i = P(X \sim x_i); \quad r_j = P(Y \sim y_j).$$

E x e m p l e 1. Trouver l'information totale mutuelle contenue dans les systèmes X et Y dans les conditions de l'exemple 1 du § 18.4.

S o l u t i o n. A partir des données de l'exemple cité et de la table 7 de l'annexe on obtient:

$$H(X, Y) = 2{,}25; \quad H(X) = 1{,}16; \quad H(Y) = 1{,}57;$$
$$I_{Y\leftrightarrow X} = H(X) + H(Y) - H(X, Y) = 0{,}48 \text{ bit}.$$

E x e m p l e 2. Un système physique X peut se trouver dans l'un des quatre états x_1, x_2, x_3, x_4; les probabilités correspondantes sont données dans le tableau:

x_i	x_1	x_2	x_3	x_4
p_i	0,1	0,2	0,4	0,3

Les états x_1 et x_2 sont indiscernables de même que les états x_3 et x_4. La communication concernant le système X indique que ce dernier se trouve dans l'un des états x_1, x_2 ou bien dans l'un des états x_3, x_4. On reçoit une communication respective. Déterminer l'information contenue dans cette communication.

S o l u t i o n. Dans l'exemple envisagé on étudie non pas le système X, mais le système dépendant Y, se trouvant dans l'état y_1 lorsque X est dans l'un des états x_1, x_2 et dans l'état y_2, lorsque X est dans l'un des états x_3, x_4. On a:

$$r_1 = P(Y \sim y_1) = 0{,}1 + 0{,}2 = 0{,}3;$$
$$r_2 = P(Y \sim y_2) = 0{,}3 + 0{,}4 = 0{,}7.$$

L'information mutuelle, c'est-à-dire l'entropie du système dépendant sera:

$$I_{Y\leftrightarrow X} = -r_1 \log r_1 - r_2 \log r_2 = \eta(0{,}3) + \eta(0{,}7) \approx 0{,}88 \text{ bit}.$$

18.6. Information partielle sur un système contenue dans une communication sur un événement. Information partielle sur un événement contenue dans une communication sur un autre événement

Dans le paragraphe précédent nous avons étudié l'information totale (ou moyenne) sur un système X contenue dans une communication nous informant de l'état du système Y. Dans certains cas il y a intérêt d'estimer l'*information partielle* sur le système X contenue dans une communication selon laquelle le système Y se trouve dans un état y_j. Désignons cette information partielle par $I_{y_j \to X}$. Remarquons que l'information totale (ou moyenne) $I_{Y \to X}$ doit être l'espérance mathématique de l'information partielle pour tous les états possibles qui peuvent être objet d'une communication:

$$I_{Y \to X} = \sum_{j=1}^{m} r_j I_{y_j \to X}. \tag{18.6.1}$$

Nous allons écrire la formule (18.5.14) de $I_{Y \to X}$ (ou de $I_{Y \leftrightarrow X}$) d'une manière analogue à (18.6.1):

$$I_{Y \to X} = \sum_{i=1}^{m} \sum_{j=1}^{m} P_{ij} \log \frac{P_{ij}}{p_i r_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} r_j P(x_i \mid y_j) \times$$

$$\times \log \frac{r_j P(x_i \mid y_j)}{p_i r_j} = \sum_{j=1}^{m} r_j \sum_{i=1}^{n} P(x_i \mid y_j) \log \frac{P(x_i \mid y_j)}{p_i}, \tag{18.6.2}$$

ce qui permet d'obtenir par comparaison avec (18.6.1) l'expression de l'information partielle:

$$I_{y_j \to X} = \sum_{i=1}^{n} P(x_i \mid y_j) \log \frac{P(x_i \mid y_j)}{p_i}. \tag{18.6.3}$$

Prenons l'expression (18.6.3) pour définition de l'information partielle. En l'analysant nous voyons que ce n'est rien d'autre que la valeur moyenne de la grandeur

$$\log \frac{P(x_i \mid y_j)}{p_i}, \tag{18.6.4}$$

prise sur tous les états x_i, pondérée par les probabilités des valeurs x_1, x_2, . . ., x_n. Comme le système Y se trouve déjà dans l'état y_j, en prenant la moyenne des grandeurs (18.6.4) on les multiplie non pas par les probabilités p_i des états x_i, mais par les probabilités conditionnelles $P(x_i \mid y_j)$.

Ainsi l'expression de l'information partielle peut s'écrire comme une espérance mathématique:

$$I_{y_j \to X} = M_{y_j}\left[\log \frac{P(X \mid y_j)}{P(X)}\right]. \tag{18.6.5}$$

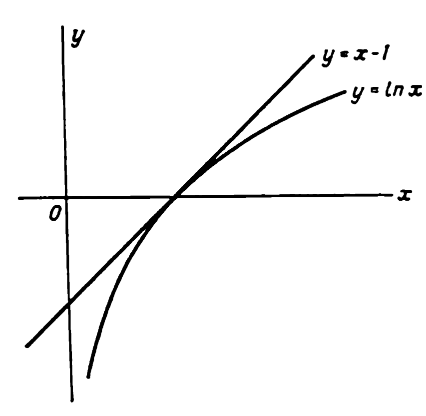


Fig. 18.6.1

Montrons que l'information partielle $I_{y_j \to X}$, tout comme l'information totale, ne peut être négative. A cet effet introduisons la désignation:

$$\frac{P(x_i \mid y_j)}{p_i} = q_{ij} \tag{18.6.6}$$

et considérons l'expression:

$$\log \frac{P(x_i \mid y_j)}{p_i} = \log q_{ij}.$$

Il est facile de voir (fig. 18.6.1) que pour tout $x > 0$ on a:

$$\ln x \leqslant x - 1. \tag{18.6.7}$$

Posant dans (18.6.7) $x = \frac{1}{q_{ij}}$ on a:

$$-\ln q_{ij} \leqslant \frac{1}{q_{ij}} - 1; \quad \ln q_{ij} \geqslant 1 - \frac{1}{q_{ij}},$$

d'où:

$$\log q_{ij} = \frac{\ln q_{ij}}{\ln 2} \geqslant \frac{1}{\ln 2}\left(1 - \frac{1}{q_{ij}}\right). \tag{18.6.8}$$

En vertu de (18.6.2) et (18.6.6) on a:

$$I_{y_j \to X} = \sum_{i=1}^{n} P(x_i \mid y_j) \log q_{ij} \geqslant \frac{1}{\ln 2} \sum_{i=1}^{n} P(x_i \mid y_j)\left(1 - \frac{1}{q_{ij}}\right) =$$

$$= \frac{1}{\ln 2} \sum_{i=1}^{n} P(x_i \mid y_j)\left[1 - \frac{p_i}{P(x_i \mid y_j)}\right] =$$

$$= \frac{1}{\ln 2}\left\{\sum_{i=1}^{n} P(x_i \mid y_j) - \sum_{i=1}^{n} p_i\right\}.$$

Mais comme

$$\sum_{i=1}^{n} P(x_i \mid y_j) = \sum_{i=1}^{n} p_i = 1,$$

l'expression dans l'accolade est nulle; par conséquent $I_{y_j \to X} \geqslant 0$.

Nous avons ainsi démontré que *l'information partielle sur le système X, contenue dans une communication relative à un état quelconque y_j du système Y ne peut être négative.* Par conséquent l'information mutuelle totale $I_{Y\leftrightarrow X}$ n'est non plus négative (étant l'espérance mathématique d'une variable aléatoire non négative):

$$I_{Y\leftrightarrow X} \geqslant 0. \qquad (18.6.9)$$

En vertu de la formule (18.5.5) donnant l'information $I_{Y\leftrightarrow X} = H(X) - H(X \mid Y)$ on a:

$$H(X) - H(X \mid Y) \geqslant 0 \qquad (18.6.10)$$

ou

$$H(X \mid Y) \leqslant H(X),$$

ce qui traduit le fait que *l'entropie conditionnelle totale d'un système n'est pas supérieure à son entropie inconditionnelle* *).

Nous avons ainsi démontré l'assertion adoptée sans démonstration dans le § 18.3.

Pour le calcul direct de l'information partielle, transformons quelque peu la formule (18.6.3), en y introduisant au lieu des probabilités conditionnelles $P(x_i \mid y_j)$ des probabilités inconditionnelles. En effet:

$$P(x_i \mid y_j) = \frac{P_{ij}}{r_j}$$

et la formule (18.6.3) devient:

$$I_{y_j \to X} = \sum_{i=1}^{n} \frac{P_{ij}}{r_j} \log \frac{P_{ij}}{p_i r_j}. \qquad (18.6.11)$$

Exemple 1. Le système (X, Y) est donné par le tableau des probabilités P_{ij}:

y_j \ x_i	x_1	x_2	r_j
y_1	0,1	0,2	0,3
y_2	0,3	0,4	0,7
p_i	0,4	0,6	

Trouver l'information partielle sur le système X contenue dans la communication $Y \sim y_1$.

*) Remarquons que ceci n'est vrai que pour l'entropie conditionnelle *totale*; pour ce qui est de l'entropie conditionnelle partielle $H(X \mid y_j)$ pour différents y_j elle peut être supérieure qu'inférieure à $H(X)$.

Solution. En vertu de la formule (18.6.11):

$$I_{y_1 \to X} = \frac{0,1}{0,3} \log \frac{0,1}{0,4 \cdot 0,3} + \frac{0,2}{0,3} \log \frac{0,2}{0,6 \cdot 0,3}.$$

La table 6 de l'annexe donne:

$$\log \frac{0,1}{0,4 \cdot 0,3} = \log 10 - \log 12 \approx -0,263;$$

$$\log \frac{0,2}{0,6 \cdot 0,3} = \log 20 - \log 18 \approx 0,152;$$

$$I_{y_1 \to X} \approx -0,333 \cdot 0,263 + 0,667 \cdot 0,152 \approx 0,013 \text{ bit}.$$

Nous avons trouvé l'information partielle sur le système X contenue dans la communication particulière sur l'événement $Y \sim y_j$. Il est naturel de se demander est-ce que l'on peut aller plus loin et déterminer l'information partielle sur l'événement $X \sim x_i$ sachant que $Y \sim y_i$? Il se trouve que ce soit possible, seulement l'information ainsi obtenue « d'un événement à l'autre » aura des particularités inattendues: elle peut être tant positive que négative.

Partant de la structure de la formule (18.6.3) on peut trouver l'information « d'un événement à l'autre » comme suit:

$$I_{y_j \to x_i} = \log \frac{P(x_i \mid y_j)}{p_i}, \tag{18.6.12}$$

donc *l'information partielle sur un événement puisée dans une communication concernant l'autre événement, est égale au logarithme du rapport de la probabilité du premier événement après la réception de la communication sur le second à la probabilité a priori de ce même événement.*

En vertu de la formule (18.6.12) si la probabilité de l'événement $X \sim x_i$ augmente après la réception de la communication $Y \sim y_j$, c'est-à-dire si

$$P(x_i \mid y_j) > p_i,$$

l'information $I_{y_j \to x_i}$ est positive; dans le cas contraire elle est négative. En particulier, lorsque la réalisation de l'événement $Y \sim y_j$ exclut complètement l'éventualité de l'événement $X \sim x_i$ (c'est-à-dire lorsque ces événements sont incompatibles), on a:

$$I_{y_j \to x_i} = -\infty.$$

L'information $I_{y_j \to x_i}$ peut s'écrire comme suit:

$$I_{y_j \to x_i} = \log \frac{P(x_i \mid y_j)}{p_i} = \log \frac{p_{ij}}{p_i r_j}, \tag{18.6.13}$$

ce qui montre qu'il y a symétrie entre x_i et y_j et donc:

$$I_{y_j \to x_i} = I_{x_i \to y_j} = I_{y_j \leftrightarrow x_i}. \tag{18.6.14}$$

Nous avons ainsi introduit trois formes d'information:

1) l'information totale sur le système X contenue dans le système Y:

$$I_{Y\to X} = I_{Y\leftrightarrow X};$$

2) l'information partielle sur le système X contenue dans l'événement (communication) $Y \sim y_j$:

$$I_{y_j\to X} \geqslant 0;$$

3) l'information partielle sur l'événement $X \sim x_i$ contenue dans l'événement (communication) $Y \sim y_j$:

$$I_{y_j\to x_i} = I_{y_j\leftrightarrow x_i}.$$

Les deux premières formes d'information ne sont pas négatives; la dernière peut être tant positive que négative.

Exemple 2. Une urne contient 3 boules blanches et 4 noires. Quatre boules ont été tirées de l'urne, trois noires et une blanche. Déterminer l'information contenue dans l'événement observé B sur l'événement A que le tirage suivant dans l'urne amènera une boule noire.

Solution. On a:

$$I_{B\to A} = \log \frac{P(A \mid B)}{P(A)} = \log \frac{1/3}{4/7} \approx 0,779 \text{ bit}.$$

18.7. Entropie et information des systèmes continus

Jusqu'à présent nous avons envisagé des systèmes physiques dont les différents états $x_1, x_2, \ldots, x_n$ étaient dénombrables; les probabilités de ces états étaient des grandeurs différentes de zéro $p_1, p_2, \ldots, p_n$. Ces systèmes sont analogues à des variables aléatoires discontinues (discrètes) prenant les valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités $p_1, p_2, \ldots, p_n$. Dans la pratique on rencontre souvent des systèmes physiques d'un type différent, analogues à des variables aléatoires continues. Les états de ces systèmes ne peuvent être numérotés: ils se transforment les uns en les autres d'une manière continue, chacun d'eux ayant une probabilité égale à zéro, la répartition des probabilités étant caractérisée par une certaine densité. Par analogie avec les variables aléatoires continues nous appellerons ces systèmes « continus », à la différence de ceux étudiés précédemment que nous appellerons « discrets ». Le système continu le plus simple est celui dont l'état est décrit par une variable aléatoire continue X de densité de probabilité $f(x)$. Dans des cas plus compliqués l'état du système est décrit par plusieurs variables aléatoires $X_1, X_2, \ldots, X_s$ de densité de probabilité $f(x_1, x_2, \ldots, x_s)$. On peut alors le considérer comme la réunion $(X_1, X_2, \ldots, X_s)$ de systèmes simples $X_1, X_2, \ldots, X_s$.

Soit un système simple X déterminé par une variable aléatoire continue X de densité de probabilité $f(x)$ (fig. 18.7.1). Essayons d'étendre à ce système la notion d'entropie introduite au § 18.1.

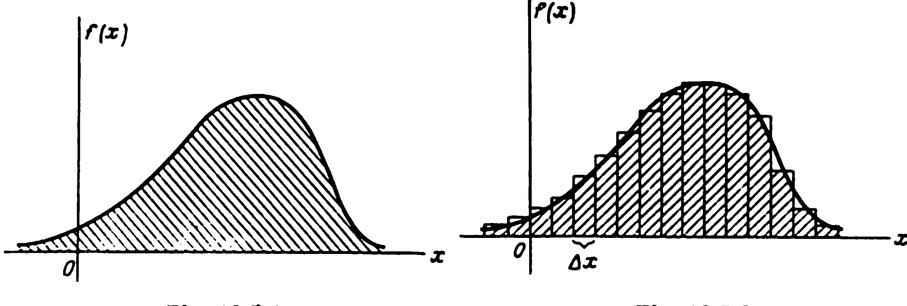


Fig. 18.7.1 Fig. 18.7.2

Il y a lieu avant tout de noter que la notion de « système continu », tout comme celle de « variable aléatoire continue », est une certaine idéalisation. Par exemple, lorsque nous considérons la variable X, par exemple, la taille d'un homme, choisi au hasard, comme une variable aléatoire continue, nous faisons abstraction du fait qu'en réalité personne ne mesure la taille avec une précision meilleure que 1 cm, et qu'il est impossible en pratique de distinguer entre elles deux tailles différant de 1 mm par exemple. Néanmoins on peut décrire cette variable aléatoire comme une variable discrète, en supposant indiscernables les valeurs différant entre elles de moins de 1 cm.

D'une manière analogue on peut, ayant établi la précision limite des mesures, c'est-à-dire ayant choisi un certain segment Δx à l'intérieur duquel les états du système X sont pratiquement indiscernables, réduire le système continu à un système discret. Ceci équivaut à remplacer la courbe continue $f(x)$ par une courbe en escalier, du type de l'histogramme (fig. 18.7.2); ici chaque segment de longueur Δx se trouve remplacé par un point représentatif. Les aires $f(x_i)\Delta x$ des rectangles représentent les probabilités de tomber sur les segments correspondants. Si l'on considère comme indiscernables des états du système associés à un même segment et que l'on les fait représenter par un même état, on peut approximativement déterminer l'entropie du système X à Δx près, on a :

$$H_{\Delta x}(X) = -\sum_i f(x_i)\,\Delta x \log\{f(x_i)\,\Delta x\} =$$

$$= -\sum_i f(x_i)\,\Delta x\,\{\log f(x_i) + \log \Delta x\} =$$

$$= -\sum_i \{f(x_i)\log f(x_i)\}\,\Delta x - \log \Delta x \sum_i f(x_i)\,\Delta x. \qquad (18.7.1)$$

Pour Δx suffisamment petit on a:

$$\sum_i \{f(x_i)\log f(x_i)\}\Delta x \approx \int_{-\infty}^{\infty} f(x)\log f(x)\,dx,$$

$$\sum_i f(x_i)\,\Delta x \approx \int_{-\infty}^{\infty} f(x)\,dx = 1$$

et la formule (18.7.1) devient alors:

$$H_{\Delta x}(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)\,dx - \log \Delta x. \qquad (18.7.2)$$

Notons que dans l'expression (18.7.2) le premier terme ne dépend pas de Δx, qui est la précision avec laquelle on détermine l'état du système. Seul le second terme ($-\log \Delta x$) en dépend, tendant vers l'infini pour $\Delta x \to 0$. C'est tout naturel, car plus on veut de précision dans la détermination de l'état du système X, plus est l'incertitude qu'il faut lever et lorsque Δx diminue indéfiniment cette incertitude croît dans la même raison.

Ainsi, si l'on se donne une zone Δx aussi petite que l'on veut « d'insensibilité » des instruments de mesure servant à déterminer l'état du système physique X, on peut trouver l'entropie $H_{\Delta x}(X)$ par la formule (18.7.2) dans laquelle le second terme croît indéfiniment lorsque $\Delta x \to 0$. Entre l'entropie $H_{\Delta x}(X)$ et ce terme croissant indéfiniment il y a une différence qui se mesure par la grandeur indépendante de Δx:

$$H^*(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)\,dx. \qquad (18.7.3)$$

Cette grandeur peut être appelée *entropie réduite* du système continu X. L'entropie $H_{\Delta x}(X)$ s'exprime en fonction de l'entropie réduite par la formule:

$$H_{\Delta x}(X) = H^*(X) - \log \Delta x. \qquad (18.7.4)$$

On peut interpréter la relation (18.7.4) comme suit: seule l'origine des calculs de l'entropie dépend de la précision de mesure.

Dans la suite pour simplifier l'écriture nous omettrons l'indice Δx dans la notation de l'entropie et écrivons simplement $H(X)$; la présence de Δx dans le second membre indiquera toujours la précision de mesure.

La formule de l'entropie (18.7.2) admet une écriture plus simple si l'on la représente tout comme dans le cas des variables continues. comme l'espérance mathématique d'une fonction. Ecrivons tout

d'abord (18.7.2) comme suit:

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log [f(x) \Delta x] \, dx. \qquad (18.7.5)$$

Ce n'est rien d'autre que l'espérance mathématique de la fonction $-\log \{f(x) \Delta x\}$ de la variable aléatoire X de densité $f(x)$:

$$H(X) = M[-\log \{f(X) \Delta x)\}]. \qquad (18.7.6)$$

On peut écrire $H^*(X)$ d'une manière analogue, soit:

$$H^*(X) = M[-\log f(X)]. \qquad (18.7.7)$$

Passons maintenant au calcul de l'entropie conditionnelle. Soient deux systèmes continus X et Y. Dans le cas général ces systèmes sont liés. Désignons par $f(x, y)$ la densité de probabilité des états du système composé (X, Y); $f_1(x)$ la densité de probabilité du système X; $f_2(y)$ la densité de probabilité du système Y; $f(y \mid x)$ et $f(x \mid y)$ les densités de probabilité conditionnelles.

Avant tout nous allons déterminer l'*entropie conditionnelle partielle* $H(Y \mid x)$, c'est-à-dire l'entropie du système Y sachant que le système X se trouve dans l'état x. La formule sera analogue à (18.4.2) mais les probabilités conditionnelles $P(y_j|x_i)$ seront remplacées par les lois de répartition conditionnelles $f(y \mid x)$ et on verra apparaître le terme $\log \Delta y$:

$$H(Y \mid x) = -\int_{-\infty}^{\infty} f(y \mid x) \log f(y \mid x) \, dy - \log \Delta y. \qquad (18.7.8)$$

Passons maintenant au calcul de l'entropie totale (moyenne) conditionnelle $H(Y \mid X)$; à cet effet il y a lieu de prendre la moyenne de l'entropie conditionnelle partielle $H(Y \mid x)$ sur tous les états x, compte tenu de leurs probabilités caractérisées par la densité $f_1(x)$:

$$H(Y \mid X) = -\iint_{-\infty}^{\infty} f_1(x) f(y \mid x) \log f(y \mid x) \, dx \, dy - \log \Delta y, \qquad (18.7.9)$$

ou compte tenu de

$$f(x, y) = f_1(x) f(y \mid x),$$

$$H(Y \mid X) = -\iint_{-\infty}^{\infty} f(x, y) \log f(y \mid x) \, dx \, dy - \log \Delta y. \qquad (18.7.10)$$

Cette formule peut s'écrire autrement:

$$H(Y \mid X) = M[-\log f(Y \mid X)] - \log \Delta y \qquad (18.7.11)$$

ou

$$H(Y \mid X) = M[-\log \{f(Y \mid X) \Delta y\}]. \qquad (18.7.12)$$

Ayant ainsi déterminé l'entropie conditionnelle nous allons montrer comment on peut l'utiliser pour trouver l'entropie d'un système composé.

Calculons tout d'abord directement l'entropie du système composé. Si les zones d'insensibilité des systèmes X et Y sont Δx et Δy, pour le système composé (X, Y) ce sera le rectangle élémentaire $\Delta x \Delta y$. On aura pour l'entropie du système (X, Y):

$$H(X, Y) = M[-\log \{f(X, Y)\, \Delta x \Delta y\}]. \qquad (18.7.13)$$

Comme

$$f(x, y) = f_1(x)\, f(y \mid x),$$

on a:

$$f(X, Y) = f_1(X)\, f(Y \mid X). \qquad (18.7.14)$$

En substituant (18.7.14) dans (18.7.13) on a:

$$H(X, Y) = M[-\log f_1(X) - \log f(Y \mid X) - \log \Delta x - \log \Delta y] =$$
$$= M[-\log (f_1(X)\, \Delta x)] + M[-\log (f(Y \mid X)\, \Delta y)],$$

ou, compte tenu des formules (18.7.6) et (18.7.12),

$$H(X, Y) = H(X) + H(Y \mid X), \qquad (18.7.15)$$

autrement dit le théorème sur l'entropie d'un système composé est applicable dans le cas des systèmes continus.

Si X et Y sont indépendants l'entropie du système composé est égale à la somme des entropies des composantes:

$$H(X, Y) = H(X) + H(Y). \qquad (18.7.16)$$

E x e m p l e 1. Trouver l'entropie du système continu X dont les états sur un certain intervalle (α, β) sont équiprobables:

$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{pour } \alpha < x < \beta, \\ 0 & \text{pour } x < \alpha \text{ ou } x > \beta. \end{cases}$$

S o l u t i o n.

$$H^*(X) = -\int_{\alpha}^{\beta} \frac{1}{\beta-\alpha} \log \frac{1}{\beta-\alpha}\, dx = \log(\beta-\alpha);$$

$$H(X) = \log(\beta-\alpha) - \log \Delta x$$

ou

$$H(X) = \log \frac{\beta-\alpha}{\Delta x}. \qquad (18.7.17)$$

E x e m p l e 2. Trouver l'entropie du système X dont les états sont répartis suivant la loi normale:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

S o l u t i o n.

$$H^*(X) = M[-\log f(X)] = M\left[-\log\left\{\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{X^2}{2\sigma^2}}\right\}\right] =$$

$$= M\left[\log(\sqrt{2\pi}\,\sigma) + \frac{X^2}{2\sigma^2}\log e\right] = \log(\sqrt{2\pi}\,\sigma) + \frac{M[X^2]}{2\sigma^2}\log e.$$

Comme:

$$M[X^2] = D[X] = \sigma^2,$$

on a:

$$H^*(X) = \log(\sqrt{2\pi}\,\sigma) + \frac{1}{2}\log e = \log(\sqrt{2\pi e}\,\sigma)$$

et

$$H(X) = \log(\sqrt{2\pi e}\,\sigma) - \log\Delta x = \log\left[\frac{\sqrt{2\pi e}\,\sigma}{\Delta x}\right]. \quad (18.7.18)$$

E x e m p l e 3. La position d'un avion est caractérisée par trois variables aléatoires: la hauteur H, le module de la vitesse V et l'angle θ déterminant la direction du vol. La hauteur de l'avion est uniformément répartie sur l'intervalle (h_1, h_2); la vitesse V suit une loi normale d'espérance mathématique v_0 et d'écart quadratique moyen σ_v; l'angle θ est uniformément réparti sur l'intervalle $(0, \pi)$. Les variables H, V, θ sont indépendantes. Trouver l'entropie du système composé.

S o l u t i o n.

D'après la formule (18.7.17) de l'exemple 1 on a:

$$H(H) = \log\frac{h_2 - h_1}{\Delta h},$$

où Δh est la « zone d'insensibilité » dans la détermination de la hauteur.

L'entropie d'une variable aléatoire ne dépendant pas de son espérance mathématique, pour déterminer l'entropie en V on peut se servir de la formule (18.7.18):

$$H(V) = \log\left[\frac{\sqrt{2\pi e}\,\sigma_v}{\Delta v}\right].$$

L'entropie du système par rapport à θ est:

$$H(\theta) = \log(\pi - 0) - \log\Delta\theta = \log\frac{\pi}{\Delta\theta}.$$

Finalement on a:

$$H(H, V, \theta) = \log\frac{h_2 - h_1}{\Delta h} + \log\frac{\sqrt{2\pi e}\,\sigma_v}{\Delta v} + \log\frac{\pi}{\Delta\theta}$$

ou

$$H(H, V, \theta) = \log\left\{\frac{h_2 - h_1}{\Delta h}\cdot\frac{\sqrt{2\pi e}\,\sigma_v}{\Delta v}\cdot\frac{\pi}{\Delta\theta}\right\}. \quad (18.7.19)$$

Notons que les facteurs dans l'accolade ont la même interprétation: ils indiquent combien de zones d'insensibilité se placent dans l'intervalle sur lequel est donnée la loi de répartition de la variable aléatoire respective. Dans le cas d'une loi uniforme c'est simplement l'intervalle des valeurs possibles de la variable aléatoire; dans le cas de la loi normale cet intervalle est égal à $\sqrt{2\pi e}\,\sigma$ où σ est l'écart quadratique moyen.

Nous avons ainsi généralisé la notion d'entropie au cas des systèmes continus. On peut d'une manière analogue généraliser la notion d'information. L'incertitude liée à la présence dans l'expression de l'entropie d'un terme augmentant indéfiniment disparaît: lorsque l'information est calculée comme une différence de deux entropies, ces termes se compensent. C'est pourquoi toute sorte d'information se rapportant à des grandeurs continues ne dépend pas de la « zone d'insensibilité ».

L'expression de l'information totale mutuelle contenue dans deux systèmes continus X et Y sera analogue à (18.5.4) avec cette seule différence que les probabilités sont remplacées par les lois de répartition et les sommes par des intégrales:

$$I_{Y\leftrightarrow X} = \int_{-\infty}^{\infty}\int f(x, y) \log \frac{f(x, y)}{f_1(x) f_2(y)}\, dx\, dy \qquad (18.7.20)$$

ou, en introduisant le signe de l'espérance mathématique,

$$I_{Y\leftrightarrow X} = M\left[\log \frac{f(X, Y)}{f_1(X) f_2(Y)}\right]. \qquad (18.7.21)$$

L'information totale mutuelle $I_{Y\leftrightarrow X}$, tout comme dans les systèmes discrets, est une grandeur non négative ne s'annulant que lorsque les systèmes X et Y sont indépendants.

E x e m p l e 4. Sur le segment (0, 1) on marque d'une manière aléatoire deux points U et V indépendamment l'un de l'autre; chacun de ces points est uniformément réparti sur cet intervalle. Dans une expérience l'un des points

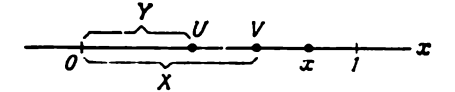


Fig. 18.7.3

s'est trouvé à droite de l'autre. Quelle est la quantité d'information sur la position du point de droite que donne la connaissance de la position du point de gauche?

S o l u t i o n. Soient deux points aléatoires U et V sur l'axe des ab cisses Ox (fig. 18.7.3). Désignons par Y l'abscisse du point de gauche, et par Xs celle du point de droite (sur la figure 18.7.3 c'est le point U qui s'est trouvé à gauche, mais l'inverse aurait pu se produire). Les variables X et Y sont associées à U et V par les relations:

$$Y = \min\{U, V\}; \quad X = \max\{U, V\}.$$

Nous allons déterminer la loi de répartition du système (X, Y). Comme $Y \leq X$ cette loi n'existera que dans le domaine D hachuré sur la figure 18.7.4. Désignons par $f(x, y)$ la densité de probabilité du système (X, Y) et calculons l'élément de probabilité $f(x, y)\, dxdy$, c'est-à-dire la probabilité pour le point aléatoire (X, Y) de se trouver dans le rectangle élémentaire $(x, x + dx;\ y,$

$y + dy$). Il y a deux éventualités possibles pour cet événement: c'est le point U qui se trouvera à gauche et le point V à droite ou bien l'inverse. Par conséquent:

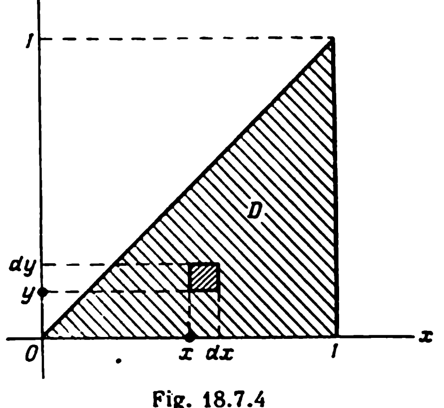


Fig. 18.7.4

$$f(x, y)\,dx\,dy = \varphi(x, y)\,dx\,dy + \varphi(y, x)\,dx\,dy,$$

où $\varphi(u, v)$ est la densité de probabilité du système de variables (U, V). Ici

$$\varphi(u, v) = 1 \quad \begin{pmatrix} 0 < u < 1 \\ 0 < v < 1 \end{pmatrix},$$

par conséquent:

$$\varphi(x, y) = \varphi(y, x) = 1;$$

$$f(x, y)\,dx\,dy = 2\,dx\,dy$$

et

$$f(x, y) = \begin{cases} 2 & \text{pour } (x, y) \subset D, \\ 0 & \text{pour } (x, y) \not\subset D. \end{cases}$$

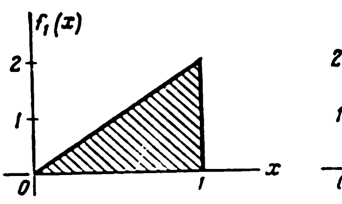
Cherchons maintenant les lois de répartition des grandeurs composant le système, on a:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^x 2dy = 2x \quad \text{pour} \quad 0 < x < 1;$$

d'une manière analogue:

$$f_2(x) = \int_{-\infty}^{\infty} f(x, y)\,dx = \int_y^1 2dx = 2(1-y) \quad \text{pour} \quad 0 < y < 1.$$

Les graphiques des densités $f_1(x)$ et $f_2(y)$ sont représentés sur la figure 18.7.5.



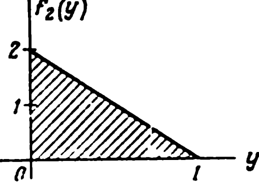


Fig. 18.7.5

En portant $f(x, y)$, $f_1(x)$ et $f_2(y)$ dans la formule (18.7.20) on obtient:

$$I_{Y \leftrightarrow X} = \frac{1}{\ln 2} \iint_{(D)} \ln \frac{1}{2x(1-y)}\, 2\,dx\,dy =$$

$$= \frac{1}{\ln 2} \left\{ -\ln 2 + \iint_{(D)} 2(-\ln x)\,dx\,dy + \iint_{(D)} 2[-\ln(1-y)\,dx\,dy \right\}.$$

Vu la symétrie du problème, les deux dernières intégrales doubles sont égales entre elles et

$$I_{Y\leftrightarrow X} = -1 - \frac{4}{\ln 2} \iint_{(D)} \ln x \, dx \, dy =$$

$$= -1 - \frac{4}{\ln 2} \int_0^1 \ln x \, dx \int_0^x dy = -1 - \frac{2}{\ln 2} \int_0^1 2x \ln x \, dx = \frac{1}{\ln 2} - 1 \approx 0{,}44 \text{ bit.}$$

E x e m p l e 5. Soit une variable aléatoire normale X, de paramètres $m_x = 0$ et σ_x. L'erreur de mesure de X est égale à Z, cette dernière suit également une loi normale de paramètres $m_z = 0$, σ_z. L'erreur Z ne dépend pas de X. On dispose du résultat de mesure

$$Y = X + Z$$

qui est également une variable aléatoire. Déterminer la quantité d'information sur X contenue dans Y.

S o l u t i o n. Utilisons pour calculer l'information la formule (18.7.21), c'est-à-dire cherchons l'information comme l'espérance mathématique de la variable aléatoire

$$U = \log \frac{f(X, Y)}{f_1(X) f_2(Y)} . \tag{18.7.22}$$

Faisons tout d'abord la transformation suivante:

$$\log \frac{f(x, y)}{f_1(x) f_2(y)} = \log \frac{f_1(x) f(y \mid x)}{f_1(x) f_2(y)} = \log \frac{f(y \mid x)}{f_2(y)} .$$

Dans notre cas

$$f_2(y) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_x^2 + \sigma_z^2}} e^{-\frac{y^2}{2(\sigma_x^2 + \sigma_z^2)}} ,$$

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\, \sigma_z} e^{-\frac{(y-x)^2}{2\sigma_z^2}} \quad \text{(voir chapitre 9).}$$

L'expression (18.7.22) vaut:

$$U = \log \frac{\sqrt{\sigma_x^2 + \sigma_z^2}}{\sigma_z} + \frac{1}{\ln 2} \left[\frac{(Y - X)^2}{2\sigma_z^2} - \frac{Y^2}{2(\sigma_x^2 + \sigma_z^2)} \right] =$$

$$= \log \frac{\sqrt{\sigma_x^2 + \sigma_z^2}}{\sigma_z} + \frac{1}{\ln 2} \left[\frac{Z^2}{2\sigma_z^2} - \frac{Y^2}{2(\sigma_x^2 + \sigma_z^2)} \right] ,$$

d'où

$$I_{Y\leftrightarrow X} = M[U] = \log \frac{\sqrt{\sigma_x^2 + \sigma_z^2}}{\sigma_z} + \frac{1}{\ln 2} \left[\frac{M[Z^2]}{2\sigma_z^2} - \frac{M[Y^2]}{2(\sigma_x^2 + \sigma_z^2)} \right] . \tag{18.7.23}$$

Mais $m_z = m_y = 0$, par conséquent:

$$\left. \begin{aligned} M[Z^2] &= D[Z] = \sigma_z^2, \\ M[Y^2] &= D[Y] = \sigma_x^2 + \sigma_z^2. \end{aligned} \right\} \tag{18.7.24}$$

En substituant (18.7.24) dans (18.7.23) on obtient:

$$I_{Y \leftrightarrow X} = \log \frac{\sqrt{\sigma_x^2 + \sigma_z^2}}{\sigma_z} \quad \text{bits *)}.$$

Par exemple, pour $\sigma_x = \sigma_z$:

$$I_{Y \leftrightarrow X} = \log \sqrt{2} = \frac{1}{2} \quad \text{bit}.$$

Si $\sigma_x = 4$; $\sigma_z = 3$, on a $I_{Y \leftrightarrow X} = \log \frac{5}{3} \approx 0{,}74$ bit.

18.8. Problèmes du codage des communications. Code de Shannon-Fano

Lorsqu'on transmet une communication on est toujours obligé à utiliser un *code* quelconque, c'est-à-dire à représenter les messages sous la forme d'une suite de signaux. Citons en exemple l'alphabet Morse adopté en télégraphie pour la transmission des communications verbales. Avec ce code on représente une communication quelconque sous la forme d'une combinaison de signaux élémentaires: point, trait, pause (espace entre lettres), pause longue (espace entre mots).

En général, on appelle *codage* la représentation de l'état d'un système physique à l'aide de l'état d'un autre système. Par exemple, dans une conversation téléphonique les signaux sonores sont codés sous la forme d'oscillations électromagnétiques, puis décodés, redevenant des signaux sonores à l'autre extrémité de la ligne. Dans le cas le plus simple de codage, les deux systèmes X et Y (à représenter et représentant) ont un nombre fini d'états possibles. Il en ait ainsi dans le cas des communications littérales, par exemple, en télégraphie. Nous allons nous limiter à ce cas simple de codage.

Soit un certain système X (par exemple les lettres de l'alphabet) pouvant prendre d'une manière aléatoire l'un des états $x_1, x_2, \ldots, x_n$. Nous voulons le coder à l'aide d'un autre système Y, dont les états possibles sont $y_1, y_2, \ldots, y_m$. Si $m < n$ on n'arrive pas à mettre en correspondance à chaque état du système X un état du système Y. On est obligé alors à représenter un état du système X par une certaine combinaison (suite) d'états du système Y. Ainsi dans l'alphabet Morse les lettres sont représentées par différentes combinaisons de symboles élémentaires (point, trait). Le choix de ces combinaisons et l'établissement d'une correspondance entre les communications à transmettre et ces combinaisons constitue codage le au sens strict du mot.

Les codes se distinguent les uns des autres par le nombre de symboles élémentaires servant à former une combinaison, en d'autres

*) On aurait pu aboutir au même résultat en utilisant la formule (18.7.20) mais par une méthode plus compliquée.

termes par le nombre d'états possibles du système Y. Dans l'alphabet Morse il y a quatre symboles élémentaires (point, trait, pause courte et pause longue). La transmission des signaux peut être réalisée sous différentes formes: signaux lumineux, émission de courant électrique de durées différentes, signaux sonores, etc.). Un code à deux symboles élémentaires (0 et 1) est appelé *binaire*. Les codes binaires sont beaucoup utilisés dans la pratique, surtout pour l'introduction de l'information dans les calculatrices électroniques fonctionnant en système binaire.

Une même communication peut être codée de différentes manières. Il se pose alors le problème du codage optimal. Il semble naturel de considérer comme optimal un code assurant la durée minimale de transmission. Si la transmission d'un symbole élémentaire (par exemple 0 ou 1) prend le même temps, le code optimal sera celui pour lequel la transmission d'une communication nécessitera un nombre minimal de symboles élémentaires.

Supposons qu'il y ait lieu de représenter en code binaire les lettres de l'alphabet français de telle sorte qu'à chaque lettre corresponde une certaine combinaison des symboles élémentaires 0 et 1 et que le nombre moyen de ces symboles par lettre du texte soit minimal.

Soit les 26 lettres de l'alphabet français: *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, *j*, *k*, *l*, *m*, *n*, *o*, *p*, *q*, *r*, *s*, *t*, *u*, *v*, *w*, *x*, *y*, *z*, plus l'intervalle entre les mots que nous désignerons par « — » et l'apostrophe « ' » (un seul symbole) ainsi que les lettres *é*, *è*, *à*, *ù* (au total 31 lettres).

La première idée est d'attribuer aux lettres, sans en changer l'ordre, un numéro de 0 à 30, puis de transformer cette numération en système binaire. Rappelons que dans le système binaire les unités des différents rangs sont des puissances de deux. Par exemple, le nombre décimal 12 peut être représenté comme

$$12 = 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$$

et donc sa forme binaire sera 1100.

Le nombre décimal 25

$$25 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

s'écrira en système binaire comme 11001.

Chacun des nombres 0, 1, 2, . . ., 30 peut être représenté par un nombre binaire à cinq chiffres. On obtient alors le code suivant:

a ~ 00000
b ~ 00001
c ~ 00010
d ~ 00011
. . . .
z ~ 11001
« — » ~ 11010

é ~ 11011
è ~ 11100
à ~ 11101
ù ~ 11110

Dans ce code il faut exactement cinq chiffres élémentaires pour représenter une lettre. La question qui se pose est de savoir si ce code est optimal et n'est-il pas possible de trouver un autre code dans lequel pour une lettre il faudra en moyenne moins de symboles élémentaires?

En effet, dans notre code pour représenter une lettre, que ce soit les lettres « a », « e », « o » que l'on rencontre fréquemment, ou les lettres « v », « w », « x », « y », « z » rares, il faut un même nombre de symboles. Il est clair qu'il serait plus raisonnable d'utiliser pour les lettres fréquentes moins de symboles que pour les lettres rares.

Pour pouvoir établir un tel code il faut évidemment connaître la fréquence d'emploi des lettres de l'alphabet français. On peut trouver ces fréquences dans le tableau 18.8.1 où les lettres sont disposées par ordre décroissant des fréquences.

Tableau 18.8.1

Lettre	Fréquence	Lettre	Fréquence	Lettre	Fréquence	Lettre	Fréquence
«—»	0,164	f	0,010	l	0,051	à	0,002
e	0,145	v	0,009	o	0,043	x	0,0004
a	0,081	b	0,009	s	0,038	z	0,0002
i	0,077	q	0,008	p	0,028	y	0,0001
n	0,063	h	0,007	d	0,028	ù	0,0001
t	0,063	è	0,005	c	0,023	w	0,0001
r	0,053	g	0,005	m	0,017	k	0,0001
u	0,053	j	0,005	é	0,012		

En utilisant ce tableau on peut établir le code le plus économique basé sur la quantité d'information. Il est évident que le code sera le plus économique lorsque chaque symbole élémentaire transmettra le maximum d'information. Considérons un symbole élémentaire (donc le signal le représentant) comme un système physique à deux états possibles: 0 et 1.

L'information associée à ce symbole est égale à l'entropie du système et est maximale dans le cas où ses deux états sont équiprobables ; le symbole élémentaire transmet alors une information qui vaut 1 bit. C'est pourquoi on met à la base du codage optimal la condition que les symboles élémentaires dans un texte codé soient en moyenne également fréquents.

Nous allons exposer ici une méthode de construction d'un code satisfaisant à cette condition ; cette méthode est connue sous le nom

de « code de Shannon-Fano ». Les symboles codés (lettres et combinaisons de lettres) sont répartis en deux groupes à peu près équiprobables : dans le premier groupe de symboles à la première place des combinaisons de code on met 0 (premier chiffre d'un nombre binaire représentant le symbole) ; dans le second groupe le nombre binaire commence par 1. Puis chaque groupe est de nouveau divisé en deux sous-groupes à peu près équiprobables ; pour les symboles du premier sous-groupe on met le zéro à la seconde place ; dans le second sous-groupe, l'unité, etc.

Nous allons illustrer l'emploi du code de Shannon-Fano sur l'exemple de l'alphabet français (tableau 18.8.1). Faisons la somme des probabilités (fréquences) des quatre premiers éléments du tableau, on obtient 0,467. La probabilité correspondant à toutes les autres lettres de « n » à « κ » est à peu près la même, soit 0,533. Les quatre premières lettres auront à la première place du code le signe binaire 0, les autres lettres, l'unité. Puis on divise de nouveau le premier groupe en deux sous-groupes à peu près équiprobables : de « — » à « *e* » et de « *a* » à « *i* » ; pour toutes les lettres du premier sous-groupe on mettra zéro à la seconde place, et pour celles du second sous-groupe, l'unité. On continue ainsi jusqu'à ce que dans chaque groupe il ne reste qu'une seule lettre qui sera codée par un certain nombre binaire. Le tableau 18.8.2 permet de comprendre le principe de formation du code, le code lui-même est donné par le tableau 18.8.3.

En se servant du tableau 18.8.3 on peut coder et décoder n'importe quelle communication. A titre d'exemple écrivons en code binaire la phrase : « théorie de l'information »

1001111111100111110011100101000100011110000100010110000111100
0111110101110010101111011010100101111001000

Notons qu'ici on n'a pas besoin de séparer les lettres les unes des autres par un signe spécial car le décodage est univoque. On peut s'en rendre compte en décodant la phrase suivante à l'aide du tableau 18.8.3 :

111101111100111100011000111100011000111101011100111100010
11111110001

(« mode de codage »).

Il faut cependant noter que toute erreur de codage (confusion accidentelle des signes 0 et 1) est dans un tel code funeste car le décodage de toute la partie du texte suivant l'erreur devient impossible. C'est pourquoi ce principe de codage peut être recommandé seulement dans les cas où les erreurs de codage et de transmission de la communication sont pratiquement exclues.

On se pose alors la question de savoir si le code établi est, en l'absence d'erreur, optimal. Pour répondre à cette question nous allons

Tableau 18.8.2

Tableau 18.8.3

Lettre	Nombre binaire	Lettre	Nombre binaire	Lettre	Nombre binaire
«—», «'»	000	p	111011	è	111111101
e	001	d	111100	g	111111110
a	010	c	1111010	j	1111111110
i	011	m	1111011	à	11111111110
n	1000	é	1111100	x	1111111111100
t	1001	f	11111010	z	1111111111101
r	1010	v	11111011	y	11111111111100
u	10110	b	11111100	u	11111111111101
l	10111	q	11111101	w	11111111111110
o	11100	h	111111100	k	11111111111111
s	111010				

trouver l'information moyenne rapportée à un symbole élémentaire (0 et 1) et la comparer à l'information maximale possible qui est égale à un bit. A cet effet trouvons tout d'abord l'information moyenne contenue dans une lettre H (l) du texte transmis, c'est-à-dire l'entropie d'une lettre:

$$H(1) = -\sum_{i=1}^{31} p_i \log p_i = \sum_{i=1}^{31} \eta(p_i),$$

p_i étant la probabilité pour que la lettre prenne un certain état (« — », a, b, c, . . ., z, é, è, à, ù).

A partir du tableau 18.8.1 on a:

$$H(1) = \eta(0,164) + \eta(0,145) + \ldots + \eta(0,0001) + \\ + \eta(0,0001) \approx 4,11$$

bits par lettre du texte.

A partir du tableau 18.8.2 on trouve le nombre moyen de symboles élémentaires par lettre:

$$n_{\text{moy}} = 3 \cdot 0,164 + 3 \cdot 0,145 + \ldots + 14 \cdot 0,0001 + \\ + 14 \cdot 0,0001 \approx 4,14.$$

En divisant l'entropie H (l) par n_{moy} on obtient l'information par symbole élémentaire:

$$I_{1s} = \frac{4,11}{4,14} = 0,993 \text{ bit}.$$

Ainsi, l'information rapportée à un symbole est très voisine de sa limite supérieure 1, donc le code adopté est très proche du code optimal. Si on reste dans les limites du *codage par lettres* on ne peut obtenir rien de mieux.

Notons que, dans le cas du codage binaire des numéros d'ordre des lettres, chaque lettre aurait été représentée par cinq symboles binaires et l'information par symbole serait:

$$I_s = \frac{4,11}{5,00} = 0,822 \text{ bit},$$

c'est-à-dire nettement inférieure à celle du codage littéral optimal.

Il faut cependant noter que le codage littéral n'est pas en général économique. En effet, il y a toujours corrélation entre les lettres d'un texte sensé quelconque. Par exemple, en français, après un « q » il y a toujours un « u »; un mot ne peut commencer par un « ù », la combinaison « pb » est impossible, etc.

Nous avons déjà vu que l'entropie d'un système composé est inférieure à la somme des entropies des composants; par conséquent, l'information transmise par un morceau de texte sensé est toujours inférieure à l'information portée par un symbole multipliée par le nombre de symboles. Compte tenu de ce qui vient d'être dit, on peut élaborer

un code plus économique en codant non pas chaque lettre séparément mais des blocs de lettres. Par exemple dans un texte français il y a lieu de coder certaines combinaisons de lettres que l'on rencontre fréquemment, par exemple « tion », « ure », « ment », etc. Les blocs codés sont disposés par ordre décroissant de fréquences comme les lettres du tableau 18.8.1, le codage binaire s'effectuant de la même manière

Dans certains cas il est raisonnable de coder non pas des blocs de lettres, mais des parties de phrases sensées. Par exemple, pour allé-

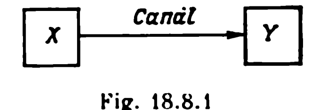


Fig. 18.8.1

ger le travail du télégraphe la veille des fêtes il est conseillé de coder par des numéros spéciaux des textes entiers comme:

« Meilleurs vœux de nouvel an ».

Sans nous arrêter spécialement sur les méthodes de codage par blocs nous nous limiterons à la formulation du théorème de Shannon.

Soit une source d'information X et un récepteur Y reliés entre eux par un canal de transmission C (fig. 18.8.1).

Le *débit* $H_1(X)$ de la source d'information est donné. Par débit en entend la quantité moyenne d'information, en bits, provenant de la source par unité de temps (numériquement égale à l'entropie de la communication débitée par la source par unité de temps). Supposons de plus que soit donnée la *capacité* C_1 du canal de transmission, c'est-à-dire la quantité maximale d'information (par exemple, de signes binaires 0 ou 1) que peut transmettre le canal en cette même unité de temps. Quelle doit être alors la capacité de transmission du canal pour qu'il puisse s'acquiter de sa tâche, donc transférer sans retard vers le récepteur Y l'information issue de la source X?

Le premier théorème de Shannon donne une réponse à cette question. Nous nous limitons à formuler ce théorème, sans en donner la démonstration.

Premier théorème de Shannon

Si la capacité de transmission C_1 d'un canal est supérieure à l'entropie de la source d'information par unité de temps

$$C_1 > H_1(X)$$

il est toujours possible de coder une communication suffisamment longue de telle sorte qu'elle soit transmise sans retard par le canal de transmission. Si au contraire

$$C_1 < H_1(X),$$

la transmission sans retard est impossible.

18.9. Transmission de l'information avec déformations. Capacité de transmission d'un canal avec bruits

Dans le paragraphe précédent nous avons envisagé les questions liées au codage et à la transmission de l'information par des canaux parfaits, lorsque la transmission de l'information se déroule sans erreurs. En fait les erreurs viennent toujours entacher l'information. Un canal de transmission où l'information subit des déformations est appelé *canal avec bruits*. Dans le cas particulier, où des erreurs apparaissent lors du codage, le dispositif de codage peut être considéré comme un canal avec bruits.

Il est évident que les bruits entraînent une perte d'information. Pour parer à cet inconvénient il y a lieu de prendre des mesures spéciales. L'une des mesures est la *redondance* de l'information transmise. On y arrive en faisant débiter la source d'information beaucoup plus de symboles qu'il n'en fallait en l'absence de bruits. La forme la plus simple d'introduction de redondance est la répétition du message. On utilise cette méthode si l'audibilité téléphonique est mauvaise. en répétant deux fois chaque message. Une autre méthode bien connue pour augmenter la fiabilité de la transmission est d'épeler le mot en transmettant au lieu de chaque lettre des mots bien connus (noms) commençant par cette lettre.

Remarquons que toutes les langues vivantes ont une certaine redondance, qui d'ailleurs permet souvent de rétablir le texte exact d'après « le sens » de la communication. C'est pourquoi les erreurs commises dans un télégramme conduisent rarement à une perte d'information : généralement on arrive à les corriger en se guidant seulement des connaissances de la langue. Ceci aurait été impossible sans redondance d'information. Pour mesurer la redondance d'une langue on se sert de la grandeur

$$U = 1 - \frac{H_{18}}{\log n} \tag{18.9.1}$$

où H_{18} est l'entropie moyenne réelle, rapportée à un symbole transmis (à une lettre), calculée pour des morceaux de texte suffisamment longs, compte tenu de la corrélation entre les symboles ; n le nombre de symboles utilisés (lettres) ; $\log n$ l'entropie maximale possible par symbole transmis, dans des conditions données, supposant que tous les symboles soient équiprobables et indépendants.

Les calculs effectués pour les langues européennes les plus répandues montrent que leur redondance atteint 50 % et même plus, ceci signifie que, grossièrement, 50 % des symboles transmis sont superflus et auraient pu ne pas être transmis, s'il n'y avait pas de danger de déformations.

Cependant pour une transmission correcte de données la redondance naturelle d'une langue peut se trouver être tant excessive qu'insuf-

fisante : tout dépend du danger de déformations (« niveau de bruits ») du canal de transmission.

Les méthodes de la théorie de l'information permettent de trouver pour un niveau de bruits donné la redondance exigée de la source d'information. Ces mêmes méthodes permettent d'élaborer des codes spéciaux antibruits (en particulier, les codes « autocorrecteurs »). Pour résoudre ces problèmes il faut savoir tenir compte de la perte d'information éventuelle dans un canal en présence de bruits.

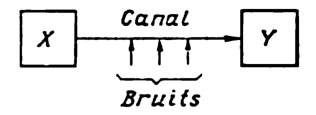


Fig. 18.9.1

Considérons un système composé, formé par une source d'information X, un canal de transmission C et d'un récepteur Y (fig. 18.9.1). La source d'information est un système physique X ayant n états possibles

$$x_1, x_2, \ldots, x_n$$

de probabilités respectives

$$p_1, p_2, \ldots, p_n.$$

Nous allons envisager ces états comme des symboles élémentaires que peut éventuellement transmettre la source X par le canal C au récepteur Y. La quantité d'information par symbole fournie par la source sera égale à l'entropie par symbole :

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i.$$

Si lors de la transmission l'information n'était pas entachée d'erreurs, la quantité d'information contenue dans le système Y sur X serait égale à l'entropie même du système X. En présence d'erreurs cette quantité sera plus petite, soit :

$$I_{Y \leftrightarrow X} = H(X) - H(X \mid Y)$$

Il est naturel de considérer l'entropie conditionnelle $H(X \mid Y)$ comme *perte d'information* par symbole élémentaire, due à la présence de bruits.

Sachant évaluer la perte d'information par symbole élémentaire, on peut trouver la capacité de transmission d'un canal avec bruits, c'est-à-dire la quantité maximale d'information qu'il peut transmettre en unité de temps.

Supposons qu'un canal puisse transmettre k symboles par unité de temps. En l'absence de bruits la capacité de transmission du canal serait égale à :

$$C = k \log n, \qquad (18.9.2)$$

car la quantité maximale d'information portée par un symbole vaut $\log n$, donc celle portée par k symboles est égale à $k \log n$, ceci si les symboles sont indépendants les uns des autres.

Considérons maintenant un canal avec bruits. Sa capacité de transmission est :

$$C = k \max I_{(1)\, Y \to X}, \qquad (18.9.3)$$

où $\max I_{(1)\, Y \to X}$ est l'information maximale par symbole pouvant être transmise en présence de bruits.

Il est en général difficile de déterminer cette information maximale car celle-ci dépend des déformations des symboles et des probabilités de ces déformations ; il importe de savoir si les symboles peuvent être confondus ou simplement perdus, si les déformations sont indépendantes ou liées, etc.

Pourtant pour les cas simples il est relativement facile de calculer la capacité de transmission d'un canal. Soit un canal C transmettant d'une source d'information X à un récepteur Y les symboles élémentaires 0 et 1, à raison de k symboles par unité de temps. Chaque symbole lors de la transmission peut être déformé (remplacé par son opposé) indépendamment des autres et ceci avec une probabilité égale à μ. Il y a lieu de calculer la capacité de transmission du canal.

Déterminons tout d'abord l'information maximale par symbole, susceptible d'être transmise par le canal. Supposons que la source débite les symboles 0 et 1 avec les probabilités respectives p et $1 - p$.

L'entropie de la source est donc :

$$H(X) = -p \log p - (1 - p) \log (1 - p).$$

Cherchons l'information $I_{(1)\, Y \to X}$ par symbole élémentaire :

$$I_{(1) Y \to X} = H(Y) - H(Y \mid X).$$

Pour trouver l'entropie conditionnelle totale $H(Y \mid X)$ il faut avant tout trouver les entropies conditionnelles partielles : $H(Y \mid X_1)$ (entropie du système Y à condition que le système X se trouve dans l'état x_1) et $H(Y \mid x_2)$ (entropie du système Y à condition que le système X se trouve dans l'état x_2). Calculons $H(Y \mid x_1)$; pour cela supposons que ce soit le symbole élémentaire 0 qui a été transmis. Nous allons trouver les probabilités conditionnelles pour le système Y de se trouver dans l'état $y_1 = 0$ et dans l'état $y_2 = 1$. La première est égale à la probabilité de non-confusion du signal :

$$P(y_1 \mid x_1) = 1 - \mu,$$

la seconde à la probabilité de confusion du signal :

$$P(y_2 \mid x_1) = \mu.$$

L'entropie conditionnelle $H(Y \mid x_1)$ sera :

$$H(Y \mid x_1) = -\sum_{i=1}^{2} P(y_i \mid x_1) \log P(y_i \mid x_1) =$$
$$= -(1-\mu)\log(1-\mu) - \mu \log \mu.$$

Calculons maintenant l'entropie conditionnelle du système Y pour la condition $X \sim x_2$ (un signal « 1 » a été transmis) :

$$P(y_1 \mid x_2) = \mu; \quad P(y_2 \mid x_2) = 1 - \mu,$$

d'où

$$H(Y \mid x_2) = -\mu \log \mu - (1 - \mu) \log (1 - \mu).$$

Ainsi

$$H(Y \mid x_1) = H(Y \mid x_2) = -\mu \log \mu -$$
$$- (1 - \mu) \log (1 - \mu). \qquad (18.9.4)$$

L'entropie conditionnelle $H(Y \mid X)$ s'obtient en prenant la moyenne des entropies conditionnelles $H(Y \mid x_1)$ et $H(Y \mid x_2)$, compte tenu des probabilités p et $1 - p$ des états x_1 et x_2. Les entropies conditionnelles étant égales entre elles, on a :

$$H(Y \mid X) = -\mu \log \mu - (1 - \mu) \log (1 - \mu).$$

D'où la conclusion suivante : *l'entropie conditionnelle $H(Y \mid X)$ ne dépend point des probabilités p et $1 - p$ des symboles 0 et 1 dans la communication transmise mais seulement de la probabilité d'erreur* μ.

Nous allons calculer l'information totale par symbole :

$$I_{(1)Y \to X} = H(Y) - H(Y \mid X) =$$
$$= \{-r \log r - (1-r)\log(1-r)\} - \{-\mu \log \mu - (1-\mu\}\log(1-\mu)\} =$$
$$= [\eta(r) + \eta(1-r)] - [\eta(\mu) + \eta(1-\mu)],$$

où r est la probabilité d'avoir à la sortie le symbole 0. Il est évident que pour des performances données du canal, l'information par symbole $I_{(1)\,Y \to X}$ est maximale quand $\eta(r) + \eta(1-r)$ est maximale. Cette fonction est maximale pour $r = 1/2$, c'est-à-dire lorsqu'à la réception les deux symboles sont équiprobables. Il est facile de voir que ceci correspond au cas où la source transmet les deux symboles avec la même probabilité $p = 1/2$. Pour cette même valeur $p = 1/2$ l'information par symbole est également maximale. La valeur maximale est :

$$I_{(1)Y \to X} = H(Y) - [\eta(\mu) + \eta(1-\mu)] = 1 - [\eta(\mu) + \eta(1-\mu)].$$

Par conséquent, dans notre cas

$$\max I_{(1)Y\to X} = 1 - [\eta(\mu) + \eta(1-\mu)],$$

et la capacité de transmission du canal est égale à:

$$C = k\,\{1 - [\eta(\mu) + \eta(1-\mu)]\}. \qquad (18.9.5)$$

Notons que $\eta(\mu) + \eta(1-\mu)$ n'est rien d'autre que l'entropie d'un système ayant deux états possibles avec les probabilités μ et $1-\mu$. Cette entropie caractérise la *perte d'information par symbole* liée à la présence de bruits dans le canal.

Exemple 1. Déterminer la capacité de transmission d'un canal pouvant transmettre 100 symboles 0 et 1 en unité de temps, chacun des symboles étant déformé (remplacé par son opposé) avec une probabilité égale à $\mu = 0{,}01$.

Solution. Dans la table 7 de l'annexe on trouve:

$$\eta(\mu) = 0{,}0664,$$
$$\eta(1-\mu) = 0{,}0144,$$
$$\eta(\mu) + \eta(1-\mu) = 0{,}0808.$$

La perte d'information par symbole est égale à 0,0808 bit. La capacité de transmission cherchée du canal est:

$$C = 100\,(1 - 0{,}0808) = 91{,}92 \approx 92 \text{ bits en unité de temps.}$$

Des calculs analogues permettent de trouver la capacité de transmission d'un canal dans des cas plus compliqués, lorsque le nombre de symboles élémentaires est supérieur à deux et les déformations des différents symboles ne sont pas indépendantes. Connaissant la capacité de transmission d'un canal, on peut trouver la limite supérieure de la vitesse de transmission de l'information par un canal avec bruits. Nous allons formuler (sans démonstration) le second théorème de Shannon relatif à ce cas.

Second théorème de Shannon

Soient une source d'information X dont l'entropie en unité de temps est égale à $\tilde{H}(X)$, *et un canal de capacité de transmission C. Si*

$$\tilde{H}(X) > C,$$

quel que soit le codage il ne peut y avoir transmission sans retard ni déformations. Au contraire, si

$$\tilde{H}(x) < C,$$

on peut toujours coder une communication suffisamment longue de telle sorte qu'elle soit transmise sans retard ni déformations avec une probabilité aussi voisine de l'unité que l'on veut.

Exemple 2. Soient une source d'information d'entropie $\tilde{H}(X) = 100$ bits en unité de temps et deux canaux de transmission de capacité 70 sym-

boles binaires (0 et 1) en unité de temps; chaque symbole binaire est remplacé par son opposé avec une probabilité égale à $\mu = 0,1$. Il faut déterminer si la capacité de transmission de ces canaux est suffisante pour transmettre l'information débitée par la source.

S o l u t i o n. Cherchons la perte d'information par symbole:

$$\eta(\mu) + \eta(1 - \mu) = 0,332 + 0,137 = 0,469 \text{ bit.}$$

La quantité maximale d'information transmise par un canal en unité de temps est:

$$C = 70\,(1 - 0,469) = 37,2 \text{ bits.}$$

La quantité maximale d'information pouvant être transmise par les deux canaux en unité de temps vaut:

$$37,2 \cdot 2 = 74,4 \text{ bits.}$$

CHAPITRE 19

ÉLÉMENTS DE LA THÉORIE DES PHÉNOMÈNES D'ATTENTE

19.1. Objet de la théorie des phénomènes d'attente

Ces dernières décennies les problèmes particuliers, ceux des phénomènes d'attente, se posent dans les domaines d'activité les plus divers. A titre d'exemple on peut citer les stations téléphoniques, les ateliers de réparation, les guichets de billets, les agences de renseignements, les coiffeurs, etc. La structure des phénomènes d'attente comprend un certain nombre d'unités de service que nous appellerons *canaux de service* ou *stations*. Les canaux de service peuvent être des lignes de communication, des employés effectuant telle ou telle opération, différents appareils, etc. Les systèmes d'attente peuvent être mono- ou multicanaux.

Tout système d'attente doit répondre aux flux de *demandes* ou d'*unités* qu'il reçoit. Les demandes arrivent l'une après l'autre à des instants qui sont dans le cas général aléatoires. L'exécution d'une demande nécessite du temps, puis le canal se trouve libre et peut exécuter la demande suivante. Tout système de service est caractérisé par la capacité d'exécution ou le débit. La théorie des phénomènes d'attente a pour objet d'établir une relation entre le caractère du flux de demandes, le rendement d'un canal et l'efficacité du service. Pour caractériser l'efficacité du service on peut, suivant les conditions du problème et les buts des recherches, utiliser différentes grandeurs et fonctions, par exemple, le pourcentage moyen de demandes refusées et quittant le système non servies ; le temps moyen d'attente des canaux et du système en entier ; le temps moyen d'attente dans la file ; la probabilité pour qu'une demande soit immédiatement servie ; la loi de répartition de la longueur des files d'attente, etc. Chacune de ces caractéristiques décrit, d'un côté ou de l'autre, l'aptitude du système à desservir le flux de demandes, ou, en d'autres termes, le débit du système.

Dans le sens strict on entend par débit le nombre moyen de demandes que le système est en mesure de desservir par unité de temps. Souvent on introduit le débit relatif, c'est-à-dire la proportion moyenne de demandes servies dans le nombre de demandes reçues. Le débit (tant absolu que relatif) dépend, dans le cas général, non seulement des paramètres du système de service mais également du caractère

du flux de demandes. Si les demandes arrivaient régulièrement à des intervalles de temps déterminés et que la durée du service était tout à fait déterminée, il serait facile de calculer le débit. En fait, les arrivées sont en général aléatoires, la plupart du temps la durée du service est également aléatoire. Par conséquent, le fonctionnement du système devient irrégulier, on voit apparaître des files d'attente ou bien, quand les demandes arrivent en quantités faibles, ce sont les stations qui doivent attendre l'unité à servir. A des arrivées aléatoires des demandes se superposent les durées aléatoires de service de certaines demandes. Ainsi, le processus de fonctionnement d'un système d'attente est un *processus aléatoire*. Pour être en mesure de donner des recommandations sur l'organisation d'un système, de déterminer son débit et lui imposer certaines exigences il y a lieu d'étudier le processus aléatoire se déroulant dans le système et de donner sa description mathématique. C'est là l'objet de la théorie des phénomènes d'attente.

Notons que ces dernières années, la théorie des phénomènes d'attente ne se borne plus à des problèmes de « service » à proprement parler, mais s'étend à divers problèmes d'automatisation industrielle où des flux de pièces à usiner peuvent être considérés comme des flux de demandes dont le rythme d'arrivée se trouve aléatoirement perturbé. Des problèmes particuliers de la théorie des phénomènes d'attente apparaissent dans l'organisation des transports. Les problèmes posés par la fiabilité des dispositifs techniques font également appel à la théorie des phénomènes d'attente; les caractéristiques des appareils telles que la durée moyenne de fonctionnement sans panne, le nombre de pièces de rechange nécessaire, la durée des réparations, etc., peuvent être déterminées en appliquant les méthodes de la théorie des phénomènes d'attente.

Des problèmes voisins de ceux de service apparaissent également dans la technique militaire. Les stations de guidage automatique, les lignes de communication, les aérodromes, les dispositifs de collecte et de traitement de l'information, etc., sont des systèmes d'attente, ayant un certain régime de fonctionnement et un certain débit.

Il est difficile d'énumérer tous les domaines d'application de la théorie des phénomènes d'attente. Ces dernières années elle est devenue une branche de la théorie des probabilités en plein essor.

Dans le présent chapitre nous allons exposer les notions élémentaires fondamentales de la théorie des phénomènes d'attente indispensables à tout ingénieur s'occupant de problèmes de gestion industrielle, économiques, sociaux ou militaires.

19.2. Processus aléatoire à un ensemble dénombrable d'états

Le processus aléatoire se déroulant dans un système d'attente c'est le changement aléatoire d'état du système : changent le nombre de canaux occupés, le nombre d'unités dans la file d'attente, etc. Ce processus diffère essentiellement de ceux que nous avons étudiés dans les chapitres 15 à 17. La différence consiste en ce que le système d'attente est un système physique du type *discret* à un nombre fini (ou dénombrable) d'états *), et le passage d'un état à l'autre se fait par saut, à l'instant où un événement se produit (arrivée d'une nouvelle demande, disponibilité d'un canal, disparition d'une demande de la file, etc.).

Soit un système physique X à états dénombrables :

$$x_1, x_2, \ldots, x_n, \ldots$$

A un instant quelconque t le système peut se trouver dans l'un de ces états. Désignons par $p_k(t)$ ($k = 1, 2, \ldots, n, \ldots$) la probabilité pour que, à l'instant t, le système se trouve dans l'état x_k. Il est évident que pour tout t on a :

$$\sum_k p_k(t) = 1. \qquad (19.2.1)$$

L'ensemble des probabilités $p_k(t)$ pour un t fixé caractérise la *section* donnée du processus aléatoire se déroulant dans le système. Loin d'être une caractéristique exhaustive du système (par exemple, l'ensemble des $p_k(t)$ ne donne aucune idée sur la corrélation entre les sections), il décrit le processus assez bien et se trouve être suffisant pour de nombreuses applications.

On distingue deux types de processus aléatoires à états dénombrables : les processus à temps discret et ceux à temps continu. Dans les premiers systèmes les passages d'un état à l'autre ne peuvent se produire qu'à des instants rigoureusement déterminés $t_1, t_2, \ldots$, séparés par les intervalles de temps finis. Dans les processus aléatoires à temps continu le passage du système d'un état à l'autre est possible à un instant quelconque.

A titre d'exemple d'un système discret X dans lequel se déroule un processus aléatoire à temps continu, on peut envisager un groupe de n avions attaquant le territoire de l'adversaire protégé par des chasseurs. On ne connaît pas d'avance ni le moment de la détection du groupe, ni les instants d'envol des chasseurs. Les différents états du système correspondent à un nombre différent d'avions abattus

---

*) En mathématiques on dit qu'un ensemble, fini ou infini, est dénombrable si l'on peut énumérer tous ses termes, c'est-à-dire l'écrire comme une suite $a_1, a_2, \ldots, a_n, \ldots$

dans le groupe d'assaut, soit :

x_0 — aucun avion n'est abattu ;
x_1 — un avion est abattu ;
. . . .
x_k — k avions sont abattus ;
. . . .
x_n — n avions sont abattus.

Sur la figure 19.2.1 on peut voir le schéma des états possibles ainsi que les transitions d'un état à l'autre.

Les flèches indiquent les transitions possibles du système. Par des flèches demi-circulaires on a figuré la possibilité pour le système

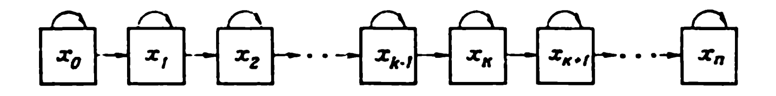


Fig. 19.2.1

de rester dans un état actuel x_k. Le système examiné se caractérise par des transitions irréversibles (les avions abattus ne peuvent être restitués) ; par suite, aucune transition à partir de l'état x_n ne se trouve être possible.

Notons que sur le schéma des transitions possibles (fig. 19.2.1) nous n'avons marqué que les transitions dans les états voisins sans figurer les sauts par plusieurs états, rejetés comme pratiquement impossibles. En effet, pour que le système saute par un état quelconque il faut que deux avions ou plus aient été abattus simultanément, mais la probabilité de cet événement est pratiquement nulle.

Les processus aléatoires se déroulant dans les systèmes d'attente sont en général des processus à temps continu. Ceci provient de ce que le flux de demandes est aléatoire. Au contraire du système à transitions irréversibles étudié dans l'exemple précédent, dans les systèmes d'attente on a des *transitions réversibles* : un canal peut devenir disponible, la file peut disparaître.

A titre d'exemple on peut considérer un système d'attente monocanal (par exemple, une ligne téléphonique) dans lequel une demande (appel) arrivant dans un canal occupé ne fait pas la queue mais quitte le système (est « refusée »). On a là un système discret à temps continu et à deux états possibles :

x_0 — le canal est libre ;
x_1 — le canal est occupé.

Les transitions d'un état à l'autre sont ici réversibles. Sur la figure 19.2.2. on a représenté les transitions possibles.

Le schéma des transitions possibles pour un système de ce genre à n canaux est donné sur la figure 19.2.3. Dans l'état x_0 tous les canaux sont libres ; dans l'état x_1 un canal est occupé, dans l'état x_2 deux canaux sont occupés, etc.

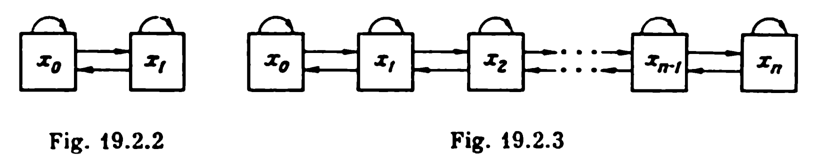


Fig. 19.2.2 Fig. 19.2.3

Soit encore un exemple du système discret à temps continu, à savoir un système d'attente monocanal pouvant se trouver dans l'un des quatre états suivants:

x_0 — le canal est en bon état et libre ;
x_1 — le canal est en bon état et occupé ;
x_2 — le canal est en panne et attend d'être réparé ;
x_3 — le canal est en panne et est en état de réparation.

Le schéma des transitions possibles pour ce cas est donné sur la figure 19.2.4 *). La transition de l'état x_3 directement dans x_1, sans passer par x_0, est pratiquement impossible, car pour cela il faudrait que la fin de la réparation coïncide exactement avec l'arrivée d'une demande.

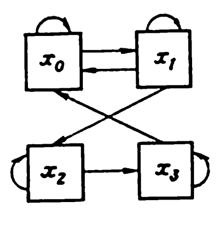


Fig. 19.2.4

Pour décrire le processus aléatoire se déroulant dans un système discret à temps continu il faut avant tout analyser les causes provoquant la transition du système d'un état à l'autre. Les processus dans les systèmes d'attente sont essentiellement déterminés par le flux de demandes. Ainsi, la description mathématique de tout système d'attente commence par la description du flux de demandes.

19.3. Flux d'événements. Flux simple et ses propriétés

En théorie des probabilités on entend par *flux d'événements* une succession d'événements à des intervalles aléatoires ou déterminés. A titre d'exemple on peut citer le flux d'appels d'une centrale téléphonique ; le flux de branchements des appareils consommateurs du courant électrique sur un réseau ; le flux de lettres recommandées arrivant à un bureau de poste ; le flux de défauts d'un ordinateur ;

---

*) Le schéma est établi en supposant qu'un canal qui à l'instant donné n'est pas en service ne peut pas tomber en panne.

le flux de coups lors d'une opération militaire, etc. Les événements formant un flux peuvent dans le cas général être très différents, mais ici nous n'allons envisager qu'un flux d'*événements homogènes*. On peut représenter un tel flux comme une suite de points t_1, t_2, . . ., t_k, . . . sur l'axe numérique (fig. 19.3.1) correspondant aux instants d'apparition des événements.

Un flux d'événements est dit *régulier* si les événements se suivent à des intervalles de temps déterminés. Dans les systèmes réels on

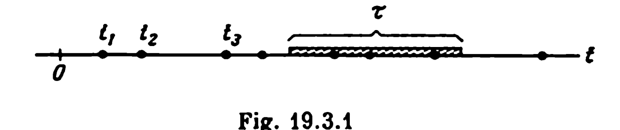


Fig. 19.3.1

rencontre rarement des flux de ce genre, cependant ils présentent un certain intérêt en tant que cas limite. Pour les systèmes d'attente plus typiques sont des flux aléatoires de demandes.

Dans le présent paragraphe nous allons envisager des flux d'événements doués des propriétés particulièrement simples. A cet effet introduisons tout d'abord les définitions nécessaires.

1. Un flux d'événements est dit *stationnaire*, si la probabilité pour un nombre quelconque d'événements de se produire dans un intervalle de temps τ (fig. 19.3.1) ne dépend que de la longueur de l'intervalle et ne dépend pas de sa position sur l'axe Ot.

2. Un flux d'événements est dit *sans postaction* si pour tous intervalles disjoints le nombre d'événements se produisant dans l'un d'eux ne dépend pas du nombre d'événements se produisant dans les autres.

3. Un flux d'événements est dit *ordinaire* si la probabilité pour deux ou plus événements de se produire dans l'intervalle élémentaire Δt est négligeable devant la probabilité correspondante d'un seul événement.

Un flux d'événements doué de trois propriétés énumérées ci-dessus est appelé *flux simple*, ou *flux stationnaire de Poisson*. Il est dit de Poisson car si les conditions 1 à 3 sont remplies, le nombre d'événements se produisant dans un intervalle fixe quelconque suit une loi de Poisson (voir § 5.9).

Nous allons étudier plus en détail les conditions 1 à 3 pour voir ce qu'elles signifient pour un flux de demandes et ce qui peut les influencer.

1. La condition de stationnarité se trouve remplie pour un flux de demandes dont les caractéristiques probabilistes ne dépendent pas du temps. En particulier, un flux stationnaire a une densité constante (nombre moyen de demandes par unité de temps). En pratique on rencontre souvent des flux de demandes qui, tout au moins sur des

intervalles de temps limités, peuvent être considérés comme stationnaires. Par exemple le flux d'appels que reçoit une station téléphonique urbaine entre 12 et 13 heures est pratiquement stationnaire. Ce même flux durant toute la journée ne saurait être stationnaire (la densité des appels de nuit est bien moindre que celle du jour). Notons qu'il en est de même pour tous les processus physiques que nous appelons stationnaires : en réalité ceux-ci ne sont stationnaires que sur des intervalles de temps limités, et si on écarte les limites vers l'infini c'est pour la commodité d'analyse. Dans de nombreux problèmes de la théorie des phénomènes d'attente il est intéressant d'étudier le fonctionnement du système pour des conditions constantes ; à cet effet il y a lieu de résoudre le problème pour un flux stationnaire de demandes.

2. La condition d'absence de postaction, condition la plus importante pour un flux simple, signifie que les demandes arrivent indépendamment les unes des autres. Par exemple, le nombre de voyageurs entrant dans une station de métro est un flux sans postaction parce que les causes qui y ont amené un quelconque voyageur sont en général indépendantes de celles des autres voyageurs. Cependant, si une dépendance apparaît, l'absence de postaction n'est plus assurée. Par exemple, le flux de passagers quittant une station de métro ne peut être considéré comme un flux sans postaction, car les instants de sortie des voyageurs arrivés par la même rame sont interdépendants.

Il faut remarquer qu'en général le flux d'unités de sortie (ou le flux de demandes servies) quittant le système d'attente a une postaction, même si le flux d'entrée n'en avait pas. Pour s'en rendre compte considérons un système d'attente monocanal pour lequel le temps de service d'une demande est déterminé et égal à t_{ser}. Dans le flux de demandes servies, l'intervalle minimum séparant les unités quittant le système sera t_{ser}. Il est facile de voir que l'existence de cet intervalle minimum conduit inévitablement à une postaction. En effet, supposons que l'on sache qu'à un certain instant t_1 une unité servie a quitté le système. On peut alors affirmer en toute certitude que dans tout intervalle de temps τ compris entre t_1 et $t_1 + t_{ser}$ aucune unité ne sera servie, ainsi il y aura une dépendance entre les nombres d'événements se produisant dans des intervalles disjoints.

On doit tenir compte du phénomène de postaction inhérent au flux de sortie si ce flux arrive à l'entrée d'un autre système d'attente (service en cascade, lorsqu'une même demande passe successivement de système en système).

Notons en passant qu'un flux régulier dans lequel les événements sont séparés par des intervalles égaux et qui semble à première vue être le flux le plus simple ne l'est pas dans le sens que nous avons adopté car il y a une postaction nette : les moments d'apparition des événements dans ce flux sont liés entre eux par une relation fonctionnelle. La postaction fait que l'analyse des processus se déroulant dans un

système d'attente pour un flux régulier de demandes est bien plus compliquée que dans le cas simple.

3. La condition d'ordinarité signifie que les demandes arrivent une à une et non par deux, trois, etc. Par exemple, le flux de clients entrant chez un coiffeur est pratiquement ordinaire, il n'est pas de même du flux de jeunes mariés arrivant à la Mairie.

Si dans un flux non ordinaire les demandes arrivent seulement par deux, par trois, etc., un tel flux se réduit facilement à un flux ordinaire, en considérant à la place du flux d'unités un flux de paires, un flux de triplets, etc. Tout se complique si chaque demande peut occasionnellement être double, triple, etc. On a alors affaire à un flux d'événements hétérogènes.

Dans la suite, pour plus de simplicité, nous nous limiterons à l'étude des flux ordinaires.

Le rôle joué par le flux simple, dans les flux d'événements, est analogue à celui de la loi normale dans les lois de répartition. On sait que lorsqu'on prend la somme d'un grand nombre de variables aléatoires indépendantes réparties suivant des lois quelconques, le résultat suit approximativement une loi normale. D'une manière analogue on peut montrer que la somme (la superposition) d'un grand nombre de flux ordinaires, stationnaires, à postaction quelconque, est un flux très voisin du flux simple. Les conditions à observer sont analogues à celles du théorème central limite, à savoir les flux s'ajoutant doivent avoir sur la somme une influence uniformément petite.

Sans démontrer cette assertion ni même formuler les conditions mathématiques auxquelles doivent satisfaire les flux, nous allons l'illustrer par des raisonnements élémentaires. Soit une suite de flux indépendants F_1, F_2, . . ., F_n (fig. 19.3.2). Prendre la « somme » de ces flux consiste à rapporter les instants d'apparition des événements à un même axe Ot, comme on peut le voir sur la figure 19.3.2.

Supposons que les flux F_1, F_2, . . . aient une influence comparable sur le flux total (c'est-à-dire que leurs densités sont du même ordre de grandeur) et que leur nombre soit suffisamment grand. Supposons de plus que ces flux soient stationnaires et ordinaires, chacun d'eux pouvant avoir une postaction, et considérons le flux total

$$F = \sum_{k=1}^{n} F_k \qquad (19.3.1)$$

sur l'axe Ot (fig. 19.3.2). Il est clair que le flux F doit être stationnaire et ordinaire car chacune de ses composantes l'est, de plus, toutes les composantes sont indépendantes. Il est intuitivement clair également que lorsque le nombre de composantes augmente, la postaction dans le flux somme, même si elle était importante dans chacun des flux, doit peu à peu s'affaiblir. En effet, considérons sur l'axe Ot deux intervalles disjoints τ_1 et τ_2 (fig. 19.3.2). Les points tombés dans ces intervalles peuvent appartenir à l'un ou l'autre de ces flux,

et lorsque n augmente, la proportion des points appartenant à un même flux (et donc dépendants) doit diminuer, les autres points appartiennent à des flux différents et donc tombent dans les intervalles τ_1, τ_2 indépendamment les uns des autres. On peut évidemment s'attendre à ce qu'avec augmentation de n le flux total n'aura pas de postaction et ressemblera fort à un flux simple.

En fait, il suffit généralement d'ajouter 4 ou 5 flux pour obtenir un flux que l'on peut considérer comme un flux simple.

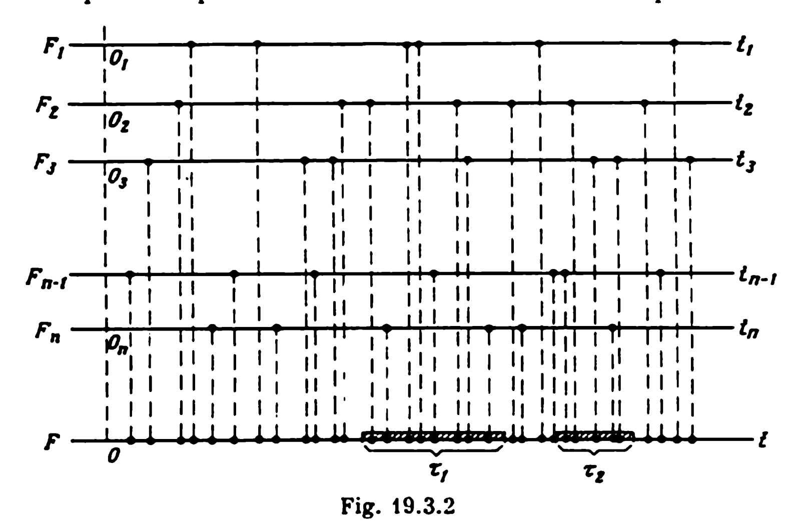


Fig. 19.3.2

Le flux simple joue dans la théorie des phénomènes d'attente un rôle particulièrement important. Premièrement, dans les applications on rencontre souvent des flux de demandes simples ou presque simples, ceci pour les raisons exposées ci-dessus. Deuxièmement, même lorsque des flux de demandes diffèrent sensiblement des flux simples on peut obtenir des résultats satisfaisants en remplaçant un flux de structure quelconque par un flux simple de même densité. C'est pour cela que nous allons étudier plus en détail un flux simple et ses propriétés.

Considérons sur l'axe Ot un flux d'événements simple F (fig. 19.3.3) comme une suite infinie de points aléatoires.

Séparons sur l'axe Ot un intervalle de longueur τ. Dans le chapitre 5 (§ 5.9) nous avons montré que, pour les conditions 1, 2 et 3 (stationnarité, absence de postaction et ordinarité) le nombre de points tombés dans l'intervalle τ suit une loi de Poisson d'espérance mathématique

$$a = \lambda\tau, \tag{19.3.2}$$

où λ est la densité du flux (nombre moyen d'événements par unité de temps).

La probabilité pour que, durant le temps τ, on ait exactement m événements est égale à :

$$P_m(\tau) = \frac{(\lambda\tau)^m}{m!} e^{-\lambda\tau}. \tag{19.3.3}$$

En particulier, la probabilité pour que l'intervalle soit vide (aucun événement n'aura lieu) sera :

$$P_0(\tau) = e^{-\lambda\tau}. \tag{19.3.4}$$

La loi de répartition de la longueur de l'intervalle séparant les événements successifs est une caractéristique importante d'un flux



Fig. 19.3.3

Considérons la variable aléatoire T, intervalle de temps entre deux événements arbitraires successifs dans un flux simple (fig. 19.3.3), et cherchons sa fonction de répartition :

$$F(t) = P(T < t).$$

Passons à la probabilité de l'événement contraire :

$$1 - F(t) = P(T \geqslant t).$$

C'est la probabilité pour que, sur un intervalle t ayant commencé à l'instant t_k d'apparition d'un des événements du flux, aucun événement postérieur n'apparaîtra. Comme un flux simple est sans post-action, la présence dans le début de l'intervalle (au point t_k) d'un événement quelconque n'influe aucunement sur la probabilité d'apparition des événements postérieurs. Ceci permet de calculer la probabilité $P(T \geqslant t)$ par la formule (19.3.4) :

$$P_0(t) = e^{-\lambda t},$$

d'où

$$F(t) = 1 - e^{-\lambda t} \quad (t > 0). \tag{19.3.5}$$

La densité de probabilité s'obtient par dérivation :

$$f(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{19.3.6}$$

La loi de répartition dont la densité s'exprime par (19.3.6) est appelée *loi exponentielle*, et la grandeur τ son *paramètre*. Le graphique de la densité $f(t)$ est donné sur la figure 19.3.4.

Comme nous allons le voir plus loin, la loi exponentielle joue un rôle très important dans la théorie des processus aléatoires discrets à temps continu. C'est pourquoi nous allor l'étudier plus en détail.

Calculons l'espérance mathématique de la variable T répartie suivant la loi exponentielle:

$$m_t = M[T] = \int_0^\infty t f(t)\,dt = \lambda \int_0^\infty t e^{-\lambda t}\,dt,$$

ou, en intégrant par parties:

$$m_t = \frac{1}{\lambda}. \qquad (19.3.7)$$

La variance de T a pour valeur:

$$D_t = D(T) = \int_0^\infty t^2 f(t)\,dt - \frac{1}{\lambda^2} = \int_0^\infty \lambda t_2 e^{-\lambda t}\,dt - \frac{1}{\lambda^2},$$

d'où

$$D_t = \frac{1}{\lambda^2}, \qquad (19.3.8)$$

$$\sigma_t = \frac{1}{\lambda}. \qquad (19.3.9)$$

Nous allons démontrer une propriété remarquable de la loi exponentielle. Elle s'énonce comme suit: *si l'intervalle de temps T réparti suivant une loi exponentielle dure déjà depuis un certain temps τ, ceci n'a aucune influence sur la loi de répartition de la partie restante de l'intervalle, celle-ci sera répartie suivant la même loi que tout l'intervalle T.*

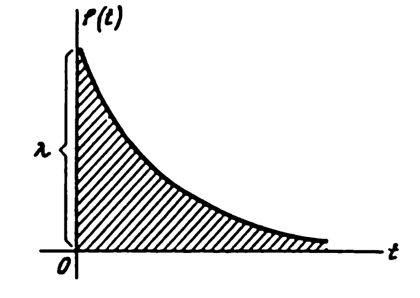


Fig. 19.3.4

Pour démontrer ceci nous considérons un intervalle de temps aléatoire T de fonction de répartition

$$F(t) = 1 - e^{-\lambda t} \qquad (19.3.10)$$

et supposons que cet intervalle dure déjà depuis un certain temps τ, c'est-à-dire que l'événement $T > \tau$ a eu lieu. Déterminons pour cette hypothèse la loi conditionnelle de répartition de la partie restante $T_1 = T - \tau$ que nous désignons $F^{(\tau)}(t)$:

$$F^{(\tau)}(t) = P(T - \tau < t \mid T > \tau). \qquad (19.3.11)$$

Montrons que la loi de répartition conditionnelle $F^{(\tau)}(t)$ ne dépend pas de τ et qu'elle est égale à $F(t)$. Pour calculer $F^{(\tau)}(t)$ trouvons tout d'abord la probabilité du produit des deux événements:

$$T > \tau \quad \text{et} \quad T - \tau < t.$$

En vertu du théorème des probabilités composées on a:

$$P((T > \tau)(T - \tau < t)) = P(T > \tau)\,P(T - \tau < t \mid T > \tau) = P(T > \tau)\,F^{(\tau)}(t),$$

d'où

$$F^{(\tau)}(t)=\frac{P((T>\tau)(T-\tau<t))}{P(T>\tau)}.$$

Mais l'événement $(T>\tau)$ $(T-\tau<t)$ est équivalent à l'événement $\tau<T<t+\tau$ dont la probabilité est:

$$P(\tau<T<t+\tau)=F(t+\tau)-F(\tau).$$

D'un autre côté:

$$P(T>\tau)=1-F(\tau),$$

par conséquent:

$$F^{(\tau)}(t)=\frac{F(t+\tau)-F(\tau)}{1-F(\tau)},$$

d'où d'après la formule (19.3.10):

$$F^{(\tau)}(t)=\frac{e^{-\lambda\tau}-e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}}=1-e^{-\lambda t}=F(t),$$

ce qu'il fallait démontrer.

Nous avons ainsi démontré que si l'intervalle de temps T est réparti suivant une loi exponentielle, l'information sur la durée antérieure de cet intervalle n'a aucune influence sur la loi de répartition de la portion restante. On peut montrer que seule la loi exponentielle est douée de cette propriété, laquelle est en fait une autre formulation de l'absence de postaction, propriété essentielle d'un flux simple.

19.4. Flux de Poisson non stationnaire

La caractéristique essentielle d'un flux d'événements non stationnaire est la *densité instantanée* $\lambda(t)$. On appelle *densité instantanée* d'un flux la limite du rapport du nombre moyen d'événements se produisant dans un intervalle de temps élémentaire $(t, t+\Delta t)$ à la longueur de cet intervalle pour $\Delta t\to 0$:

$$\lambda(t)=\lim_{\Delta t\to 0}\frac{m(t+\Delta t)-m(t)}{\Delta t}=m'(t), \qquad (19.4.1)$$

où $m(t)$ est l'espérance mathématique du nombre d'événements dans l'intervalle $(0, t)$.

Considérons un flux d'événements homogènes, ordinaire et sans postaction, mais non stationnaire, de densité variable $\lambda(t)$. Un tel flux est appelé *flux de Poisson non stationnaire*. C'est la première étape de généralisation d'un flux simple. Il est facile de montrer par une méthode analogue à celle du § 5.9 que pour un tel flux le nombre d'événements se produisant dans l'intervalle de longueur τ dont l'origine est au point t_0 suit une loi de Poisson:

$$P_m(\tau, t_0)=\frac{a^m}{m!}e^{-a} \quad (m=0, 1, 2, \ldots), \qquad (19.4.2)$$

où a est l'espérance mathématique du nombre d'événements dans l'intervalle de t_0 à $t_0 + \tau$ qui a pour valeur:

$$a = \int_{t_0}^{t_0+\tau} \lambda(t)\, dt. \qquad (19.4.3)$$

La grandeur a dépend ici non seulement de la longueur τ de l'intervalle, mais également de sa position sur l'axe Ot.

Recherchons pour un flux non stationnaire la loi de répartition de l'intervalle de temps T entre des événements voisins. Ce flux n'étant pas stationnaire, la loi cherchée dépendra de la position du premier événement sur l'axe Ot. De plus, cette loi dépendra de la forme de la fonction $\lambda(t)$. Soit t_0 l'instant d'apparition du premier des deux événements voisins; trouvons pour cette condition la loi de répartition du temps T entre cet événement et le suivant:

$$F_{t_0}(t) = P(T < t) = 1 - P(T \geqslant t).$$

La probabilité $P(T \geqslant t)$ pour que dans l'intervalle entre t_0 et $t_0 + t$ aucun événement n'apparaisse est:

$$P(T \geqslant t) = e^{-a} = e^{-\int_{t_0}^{t_0+t} \lambda(t)dt}$$

d'où

$$F_{t_0}(t) = 1 - e^{-\int_{t_0}^{t_0+t} \lambda(t)dt} \qquad (19.4.4)$$

Par dérivation on trouve la densité de probabilité:

$$f_{t_0}(t) = \lambda(t_0 + t)\, e^{-\int_{t_0}^{t_0+t} \lambda(t)dt} \quad (t > 0). \qquad (19.4.5)$$

Cette loi de répartition n'est plus exponentielle. Sa forme dépend du paramètre t_0 et de la forme de la fonction $\lambda(t)$. Par exemple, pour une variation linéaire de $\lambda(t)$

$$\lambda(t) = a + bt$$

la densité (19.4.5) s'écrit:

$$f_{t_0}(t) = [a + b(t_0 + t)]\, e^{-at - bt_0t - \frac{bt^2}{2}}. \qquad (19.4.6)$$

Le graphique de cette loi pour $a = 0{,}4$; $b = 2$ et $t_0 = 0{,}3$ est donné sur la figure 19.4.1.

Bien que la structure d'un flux de Poisson non stationnaire soit un peu plus compliquée que celle d'un flux simple, il est très commode pour les applications pratiques, car la propriété essentielle d'un flux

simple, l'absence de postaction, s'y trouve conservée. Ainsi, si l'on se donne sur l'axe Ot un point arbitraire t_0, la loi de répartition $f_{t_0}(t)$ du temps T séparant ce point du premier événement futur ne dépend

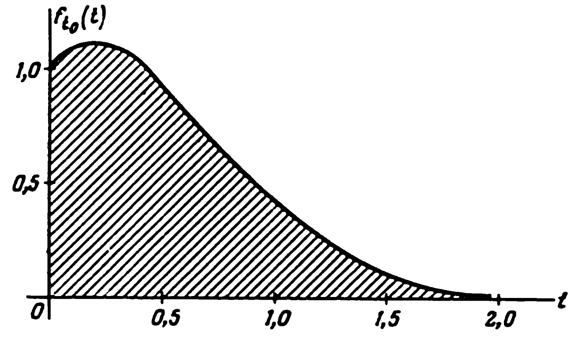


Fig. 19.4.1

pas de ce qui s'est passé avant l'instant t_0 et au point t_0, c'est-à-dire que peu importe si d'autres événements ont eu lieu antérieurement ou non.

19.5. Flux à postaction limitée (flux de Palma)

Dans le paragraphe précédent nous avons vu une généralisation naturelle du flux simple, à savoir le flux de Poisson non stationnaire. Le flux à postaction limitée en est une autre généralisation.

Considérons un flux ordinaire d'événements homogènes (fig. 19.5.1). Ce flux est appelé *flux à postaction limitée* (ou *flux de*

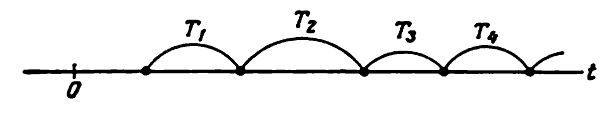


Fig. 19.5.1

Palma) si les intervalles de temps T_1, T_2, . . . entre les événements successifs sont des variables aléatoires indépendantes.

Il est clair que le flux simple est un cas particulier du flux de Palma : dans le flux simple les intervalles $T_1, T_2, \ldots$ sont des variables aléatoires indépendantes réparties suivant une loi exponentielle. Pour ce qui est du flux de Poisson non stationnaire, ce n'est pas un flux de Palma. En effet, considérons deux intervalles T_k et T_{k+1} dans un flux de Poisson non stationnaire. Comme nous l'avons vu dans le paragraphe précédent, la loi de répartition de l'intervalle entre des événements dans un flux non stationnaire dépend de l'ori-

gine de cet intervalle, le début de l'intervalle T_{k+1} coïncidant avec la fin de l'intervalle T_k, alors les longueurs de ces intervalles sont dépendantes.

Nous allons voir quelques exemples de flux de Palma.

1. Une certaine pièce d'un appareil technique (un tube électronique par exemple) fonctionne d'une manière continue jusqu'à ce qu'elle soit hors d'usage, à ce moment elle est instantanément remplacée par une autre. La durée de fonctionnement sans défaillance

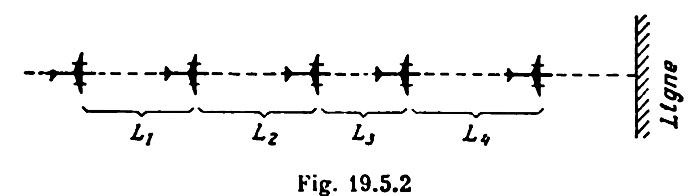


Fig. 19.5.2

est aléatoire; les défauts dans les différentes pièces surviennent indépendamment les uns des autres. Dans ces conditions le flux de défauts (ou le flux de réparations) est un flux de Palma. Si de plus la durée de fonctionnement de la pièce suit une loi exponentielle, le flux de Palma devient un flux simple.

2. Un groupe d'avions va en « colonne » (fig. 19.5.2) avec une vitesse V. Chaque avion sauf le chef de file doit se trouver à une certaine distance L du précédent. Vu l'erreur de mesure du radiotélémètre de bord, cette distance se trouve maintenue avec une certaine erreur. Les instants de traversée par les avions d'une certaine ligne forment un flux de Palma, car les variables aléatoires $T_1 = \frac{L_1}{V}$; $T_2 = \frac{L_2}{V}$; . . . sont indépendantes. Notons que ce même flux ne sera plus un flux de Palma si chaque avion tend à maintenir une distance constante non pas de l'avion précédent mais du chef de file.

Les flux de Palma sont souvent des flux de sortie des systèmes d'attente. Un flux de demandes arrivées à un système d'attente se trouve divisé par ce système en deux: un flux de demandes servies et un flux de demandes refusées. Ce dernier étant dirigé vers un autre système d'attente, il y a intérêt d'étudier ses propriétés.

Le théorème de Palma, que nous énoncerons sans le démontrer, est le théorème fondamental de la théorie des flux de sortie.

Supposons qu'un système d'attente reçoive un flux de demandes du type de Palma et qu'une demande arrivant lorsque tous les canaux sont occupés soit refusée. Si le temps de service suit une loi exponentielle, le flux des demandes refusées est également un flux de Palma.

En particulier, si le flux d'entrée est simple, le flux des demandes refusées, sans être simple, sera quand même avec une postaction limitée.

Un exemple intéressant de flux à postaction limitée est fourni par des flux d'Erlang obtenus par « tamisage » d'un flux simple.

Considérons un flux simple (fig. 19.5.3) et prenons dans ce flux un point sur deux (sur la figure les points rejetés sont marqués d'une croix). Les points restants forment un flux qui est appelé *flux d'Erlang*

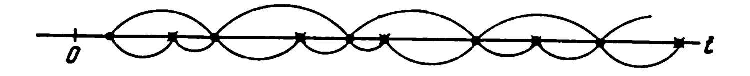


Fig. 19.5.3

du premier ordre (E_1). Ce flux est évidemment un flux de Palma, car de l'indépendance des intervalles entre les événements dans un flux simple résulte celle des grandeurs T_1, T_2, . . . qui sont des sources de deux intervalles consécutifs du flux initial.

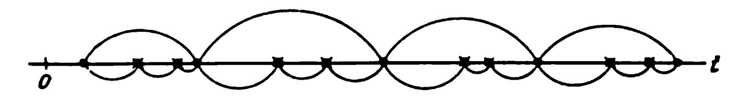


Fig. 19.5.4

Le *flux d'Erlang du second ordre* s'obtient en conservant dans un flux simple un point sur trois en laissant tomber deux points intermédiaires (fig. 19.5.4).

D'une façon générale, on appelle *flux d'Erlang d'ordre* k (E_k) le flux obtenu à partir du flux simple où l'on conserve un point sur

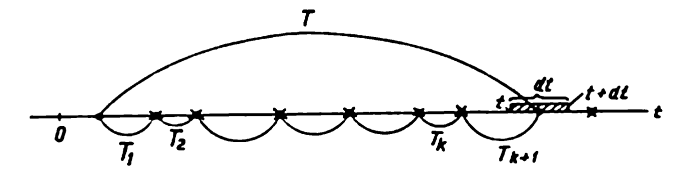


Fig. 19.5.5

tous les $(k + 1)$ points et laisse tomber les autres. Il est évident qu'on peut considérer un flux simple comme un flux d'Erlang d'ordre zéro (E_0).

Recherchons la loi de répartition de l'intervalle de temps T entre des événements voisins dans un flux d'Erlang d'ordre k (E_k). Soit sur l'axe Ot (fig. 19.5.5) un flux simple à intervalles T_1, T_2, . . . La grandeur T est une somme de $k + 1$ variables aléatoires indépen-

dantes T_i $(i = 1, 2, \ldots, k+1)$:

$$T = \sum_{i=1}^{k+1} T_i, \tag{19.5.1}$$

suivant chacune une même loi exponentielle

$$f(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{19.5.2}$$

Il serait possible de trouver la loi de répartition de T par composition de $(k+1)$ lois (19.5.2); mais il est plus facile de la déduire en raisonnant comme suit.

Désignons par $f_k(t)$ la densité de probabilité de la variable T pour le flux E_k; $f_k(t)\,dt$ étant la probabilité pour que la variable T se trouve entre t et $t + dt$ (fig. 19.5.5). Ceci signifie que le dernier point de l'intervalle T doit tomber dans l'intervalle élémentaire $(t, t + dt)$, et les k points antérieurs du flux simple dans l'intervalle $(0, t)$. La probabilité du premier événement est λdt; la probabilité du second en vertu de la formule (19.3.2) sera:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

En multipliant ces probabilités on a:

$$f_k(t)\,dt = \frac{\lambda (\lambda t)^k}{k!} e^{-\lambda t}\,dt,$$

d'où

$$f_k(t) = \frac{\lambda (\lambda t)^k}{k!} e^{-\lambda t} \quad (t > 0). \tag{19.5.3}$$

La loi de répartition donnée par la densité de probabilité (19.5.3) s'appelle *loi d'Erlang d'ordre k*. Il est évident que pour $k = 0$ elle devient une loi exponentielle, c'est-à-dire

$$f_0(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{19.5.4}$$

Recherchons les caractéristiques de la loi d'Erlang $f_k(t)$, à savoir son espérance mathématique m_k et sa variance D_k. En vertu du théorème d'addition des espérances mathématiques on a:

$$m_k = \sum_{i=1}^{k+1} m_0 = (k+1)\,m_0,$$

où $m_0 = \frac{1}{\lambda}$ est l'espérance mathématique de l'intervalle entre les événements dans un flux simple.

Il en résulte:

$$m_k = \frac{k+1}{\lambda}. \tag{19.5.5}$$

De même, le théorème d'addition des variances donne:

$$D_k = \frac{k+1}{\lambda^2}, \quad \sigma_k = \frac{\sqrt{k+1}}{\lambda}. \tag{19.5.6}$$

La densité Λ_k du flux E_k est l'inverse de m_k, soit:

$$\Lambda_k = \frac{\lambda}{k+1}. \tag{19.5.7}$$

Ainsi, lorsque l'ordre du flux d'Erlang augmente, l'espérance mathématique et la variance de l'intervalle de temps séparant les événements successifs augmentent et la densité du flux diminue.

Nous allons voir quelles modifications subit le flux d'Erlang lorsque $k \to \infty$, mais la densité reste constante. A cet effet, normons T de telle sorte que son espérance mathématique (et, par conséquent, la densité du flux) reste constante. Pour cela il nous faut changer l'échelle des temps en considérant au lieu de T la grandeur

$$\tilde{T} = \frac{T}{k+1}. \tag{19.5.8}$$

Ce flux est dit *flux d'Erlang normé d'ordre k*. La loi de répartition de l'intervalle $\tilde{T}$ entre les événements de ce flux sera:

$$\tilde{f}_k(t) = \frac{\Lambda_k (\Lambda_k t)^k}{k!} e^{-\Lambda_k t} \quad (t > 0), \tag{19.5.9}$$

où $\Lambda_k = \lambda(k+1)$, ainsi:

$$\tilde{f}_k(t) = \frac{\lambda(k+1)}{k!} [\lambda(k+1)t]^k e^{-\lambda(k+1)t} \quad (t > 0). \tag{19.5.10}$$

L'espérance mathématique de la variable $\tilde{T}$ répartie suivant la loi (19.5.10) ne dépend pas de k, elle est égale à:

$$\tilde{m}_k = m_0 = \frac{1}{\lambda},$$

où λ est la densité du flux, égale pour k quelconque à la densité du flux initial simple. La variance de $\tilde{T}$ est égale à:

$$\tilde{D}_k = \frac{D_k}{(k+1)^2} = \frac{1}{\lambda^2 (k+1)} \tag{19.5.11}$$

et décroît indéfiniment lorsque k augmente.

Ainsi on arrive à la conclusion suivante: *lorsque k augmente indéfiniment, le flux normé d'Erlang tend vers un flux régulier à intervalles constants égaux* à $1/\lambda$.

Cette propriété des flux d'Erlang est très commode pour les applications pratiques: en choisissant k convenablement on peut varier dans de larges limites la puissance de postaction, de l'absence totale ($k = 0$) jusqu'à une relation fonctionnelle déterminée entre les instants d'apparition des événements ($k = \infty$). L'ordre du flux d'Er-

lang peut donc servir de mesure de postaction d'un flux. Pour simplifier le problème, souvent il est commode de remplacer le flux réel de demandes, avec postaction, par un flux d'Erlang normé ayant approximativement les mêmes valeurs de l'espérance mathématique et de la variance de l'intervalle entre les arrivées de demandes.

E x e m p l e. Par traitement statistique des durées T des intervalles entre les demandes on a obtenu les estimations suivantes de l'espérance mathématique et de la variance de T:

$$m_t = 2 \text{ (mn)}, \qquad D_t = 0{,}8 \text{ (mn}^2\text{)}.$$

Remplacer ce flux par un flux d'Erlang normé de mêmes caractéristiques.

S o l u t i o n. On a:

$$\lambda = \frac{1}{m_t} = 0{,}5.$$

La formule (19.5.11) donne:

$$k+1 \approx \frac{1}{D_t \lambda^2} = \frac{1}{0{,}2} = 5, \qquad k = 4.$$

On peut remplacer le flux en question par un flux d'Erlang normé d'ordre 4.

19.6. Temps de service

En plus des caractéristiques du flux des arrivées, le régime de fonctionnement du système dépend également des caractéristiques du système: du nombre de canaux n et de la rapidité de fonctionnement de chacun des canaux. Une des caractéristiques essentielles du système est le *temps de service d'une demande* T_{ser}. Cette grandeur peut être tant déterminée qu'aléatoire. Il est évident que le cas d'un temps de service aléatoire est plus général.

Considérons la variable aléatoire T_{ser} et désignons par $G(t)$ sa fonction de répartition:

$$G(t) = P(T_{ser} < t) \tag{19.6.1}$$

et par $g(t)$ la densité de probabilité:

$$g(t) = G'(t). \tag{19.6.2}$$

Le plus grand intérêt pour les applications pratiques présente le cas où T_{ser} suit une loi exponentielle:

$$g(t) = \mu e^{-\mu t} \qquad (t > 0), \tag{19.6.3}$$

où le paramètre μ est une grandeur inversement proportionnelle au temps moyen de service d'une demande:

$$\mu = \frac{1}{m_{t_{ser}}}, \; m_{t_{ser}} = M[T_{ser}]. \tag{19.6.4}$$

La place particulière occupée dans la théorie des phénomènes d'attente par la loi de répartition exponentielle de T_{ser} est liée à

la propriété de cette loi démontrée au § 19.4. Appliquée au temps de service, cette propriété s'énonce : si à un instant quelconque t_0 une demande est en cours de service, la loi de répartition du temps *que durera le service* ne dépend pas du temps *écoulé dès le début du service*.

A première vue il peut paraître artificiel de supposer que le temps de service soit réparti suivant une loi exponentielle. Dans certaines applications il semble plus naturel de le supposer non aléatoire ou réparti suivant une loi normale. Cependant dans certaines conditions le temps de service suit réellement une loi voisine de l'exponentielle.

Ce sont tout d'abord les cas où le service se réduit à une série d'« essais » dont le résultat apparaît avec une probabilité égale à p.

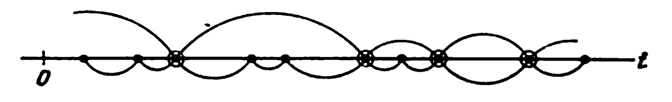


Fig. 19.6.1

Supposons que le « service » consiste à tirer dessus un objectif et se termine par sa destruction. On tire des coups indépendants avec une vitesse de λ coups par unité de temps. Chaque coup atteint le but avec une probabilité égale à p. Pour ne pas être obligé de tenir compte du moment de chaque coup on suppose que ceux-ci sont tirés à des instants aléatoires et forment un flux simple F de densité λ (fig. 19.6.1).

Séparons dans ce flux un flux de « bons » coups (marqués d'un cercle sur la figure 19.6.1). On dit qu'un coup est « bon » s'il atteint son but (à moins que le but ne fût pas atteint avant). Il est facile de voir que les « bons » coups forment également un flux simple F* de densité $\Lambda = \lambda p$ (car le flux initial F est simple et chaque coup peut atteindre le but indépendamment des autres avec une probabilité égale à p). La probabilité pour que le but soit atteint avant l'instant t sera :

$$G(t) = P(T_{\text{ser}} < t) = 1 - e^{-\Lambda t},$$

d'où la densité de probabilité du temps de « service » :

$$g(t) = \Lambda e^{-\Lambda t},$$

mais c'est l'expression d'une loi exponentielle de paramètre $\mu = \Lambda$.

On peut également utiliser d'une manière approchée la loi exponentielle. Le temps de tir avant destruction de l'objectif peut être supposé réparti suivant une loi exponentielle, bien que d'une manière approchée, également dans le cas où les coups ne forment pas un flux simple mais sont séparés par des intervalles de temps réguliers t_1, si toutefois la probabilité p de destruction par un coup n'est pas trop

grande. Pour illustrer ce qui vient d'être dit nous allons porter sur un même graphique (fig. 19.6.2) la fonction de répartition du temps de tir avant le coup destructeur (courbe en échelons) pour le cas $p = 0{,}4$; $t_1 = 1$, et la fonction de répartition de la loi exponentielle de paramètre $\mu = p = 0{,}4$ (courbe continue). On voit que la loi exponentielle traduit bien le caractère croissant de la fonction de répartition pour le cas discret. Naturellement, la correspondance est meilleure si les instants des coups ne sont pas strictement déterminés.

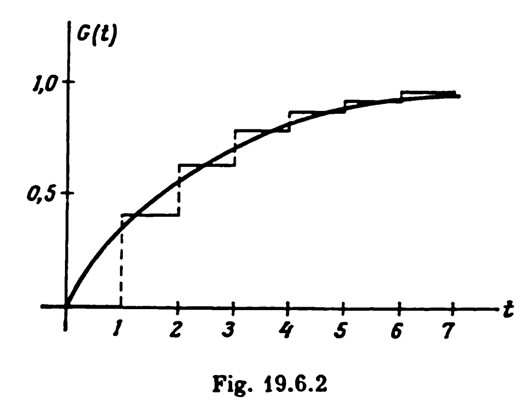


Fig. 19.6.2

L'exemple qui vient d'être envisagé n'est pas l'unique cas où le service est une série d'« essais ». On peut rapporter à ce type de service le dépistage des défauts des dispositifs techniques, lorsque la recherche des pièces défectueuses se fait en les testant l'une après l'autre. On peut y rapporter également les problèmes où le « service » est la détection d'un objectif quelconque par un radar, si l'objectif peut, avec une certaine probabilité, être détecté par chaque tour d'horizon.

La loi exponentielle décrit bien les processus où la densité de probabilité du temps de service, pour une raison ou une autre, décroît lorsque l'argument t croît. Ceci a lieu lorsque la majorité des demandes se trouvent servies très rapidement et que les retards importants de service sont très rares. Soit, par exemple, le guichet d'un bureau de poste, où l'on peut acheter des enveloppes et des timbres et d'où l'on peut envoyer des expéditions et des mandats. La majorité des clients achètent des timbres et des enveloppes et sont très rapidement servis. L'envoi de lettres recommandées est plus rare et il faut un peu plus de temps pour servir ces clients. Les mandats sont encore plus rares et leur expédition dure encore plus longtemps. Enfin parfois, mais c'est très rare, les expéditeurs envoient à la fois une grande quantité de lettres. L'histogramme de la répartition du temps de service est représenté sur la figure 19.6.3. Comme la densité de probabilité décroît lorsque t augmente, on peut, sans commettre de grande erreur,

approcher la répartition en question par la loi exponentielle, à condition de choisir convenablement son paramètre μ.

Bien entendu, la loi exponentielle n'est pas la loi universelle de répartition du temps de service. Dans beaucoup de cas c'est la loi d'Erlang qui décrit mieux le temps de service. Heureusement, le débit et certaines autres caractéristiques des systèmes d'attente dépendent peut de la formé de la loi du temps de service, mais essentiellement de sa valeur moyenne $m_{t_{\text{ser}}}$. C'est pourquoi, dans la théorie

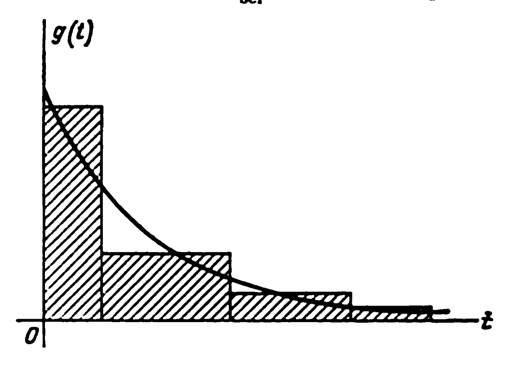


Fig. 19.6.3

des phénomènes d'attente, on suppose que le temps de service est réparti suivant la loi exponentielle. Cette hypothèse permet de simplifier l'appareil mathématique utilisé pour la solution des problèmes posés par des phénomènes d'attente et dans de nombreux cas de trouver des formules analytiques simples pour les caractéristiques du débit du système.

19.7. Processus aléatoire de Markov

Les hypothèses que le flux de demandes est poissonien et la loi de répartition du temps de service exponentielle sont d'autant plus avantageuses qu'elles permettent d'appliquer à la théorie des phénomènes d'attente la théorie des processus aléatoires de Markov.

On dit que le processus se déroulant dans un système physique est un processus de Markov (ou processus sans postaction), *si à chaque instant la probabilité d'un état quelconque du système dans le futur dépend seulement de l'état du système à l'instant actuel* (t_0) *sans dépendre de la manière dont le système a été amené dans cet état.*

Nous allons voir un exemple particulièrement simple du processus aléatoire de Markov. Un point X se déplace au hasard sur l'axe des abscisses Ox. A l'instant $t = 0$ le point X se trouve à l'origine des coordonnées ($x = 0$) et y reste durant une seconde. Au bout de cette seconde on jette une pièce de monnaie ; si l'on obtient pile, le

point X se déplace d'une unité de longueur à droite, si l'on obtient face, le déplacement est à gauche. Une seconde après on jette de nouveau une pièce de monnaie et on effectue un nouveau déplacement du point en conformité du côté apparu, etc. Les déplacements aléatoires du point forment un processus aléatoire à temps discret ($t =$ $= 0, 1, 2, \ldots$) et à états dénombrables:

$$x_0 = 0;\ x_1 = 1;\ x_{-1} = -1;\ x_2 = 2;\ x_{-2} = -2;\ldots$$

On a montré sur la figure 19.7.1 le schéma des transitions possibles pour ce processus.

Nous allons montrer que c'est un processus de Markov. En effet, supposons pour fixer les idées qu'à un certain instant t_0 l'état du sys-

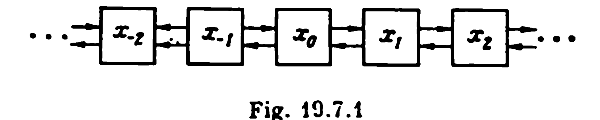


Fig. 19.7.1

tème est x_1: le point se trouve à une distance unité à droite de l'origine des coordonnées. Dans une seconde que nous prenons pour unité de temps le point se trouvera dans l'une des positions x_0 ou x_2 avec les probabilités $\frac{1}{2}$ et $\frac{1}{2}$; dans deux secondes il occupera l'une des positions x_{-1}, x_1, x_3 avec les probabilités respectives $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ et ainsi de suite. Il est évident que toutes ces probabilités dépendent de la position du point à l'instant t_0 et ne dépendent pas du tout de la manière dont il y est arrivé.

Soit un autre exemple. Un dispositif technique X se compose d'éléments (pièces) des types a et b ayant une durée de vie différente. Ces éléments peuvent se trouver défectueux indépendamment les uns des autres et à des instants aléatoires. Pour que le dispositif fonctionne il faut que tous les éléments soient en bon état. La durée de fonctionnement sans défaillance d'un élément est une variable aléatoire à répartition exponentielle; pour les éléments des types a et b les paramètres de cette loi sont différents et respectivement égaux à λ_a et λ_b. En cas de défaut du dispositif des mesures sont immédiatement prises pour en trouver les causes et l'élément défectueux est instantanément remplacé par un neuf. Le temps de réparation du dispositif est réparti suivant une loi exponentielle de paramètre μ_a si c'est un élément du type a qui s'est trouvé défectueux, et μ_b si c'est un élément du type b.

Dans cet exemple le processus aléatoire se déroulant dans le système est un processus de Markov à temps continu et à nombre fini d'états:

x_0 — aucun élément n'est défectueux, le système fonctionne;

x_1 — un élément du type a est défectueux, le système est en réparation *);

x_2 — un élément du type b est défectueux, le système est en réparation.

Sur la figure 19.7.2 on peut voir le schéma des transitions possibles.

En effet, le processus est markovien. Supposons qu'à l'instant t_0 le système se trouve dans l'état x_0 (bon état de fonctionnement). Comme la durée de fonctionnement sans défaillance de chaque élément est exponentielle, l'instant de sa défaillance future ne dépend pas du temps écoulé dès le début du fonctionnement de l'élément. Ainsi, la probabilité pour le système de rester dans l'état x_0 et d'en sortir ne dépend pas de la préhistoire du processus. Supposons maintenant qu'à l'instant t_0 le système se trouve dans l'état x_1 (un élément du type a est défectueux). Comme le temps de réparation suit également une loi exponentielle, la probabilité pour que la réparation soit finie à un instant quelconque après t_0 ne dépend pas du début de la réparation ni du moment où les autres éléments ont été montés. Le processus étudié est donc bien markovien.

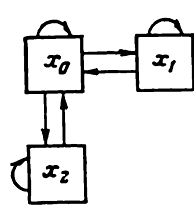


Fig. 19.7.2

Notons que la répartition exponentielle du temps de fonctionnement sans défaillance d'un élément et celle du temps de réparation sont des conditions essentielles sans lesquelles le processus ne saurait être markovien. En effet, supposons que le temps de fonctionnement sans défaillance d'un élément suive une loi autre qu'exponentielle, par exemple, la loi uniforme sur l'intervalle (t_1, t_2). Ceci signifie que chaque élément est garanti de fonctionner durant le temps t_1 et sur l'intervalle de t_1 à t_2 peut être en défaut à un instant quelconque avec une même probabilité. Supposons qu'à l'instant t_0 un élément soit en bon état. La probabilité pour que l'élément soit en défaut, sur un intervalle de temps quelconque dans le futur, dépend évidemment du temps écoulé depuis sa mise en place, c'est-à-dire de la « préhistoire », et le processus ne saurait être markovien.

Il en est de même du temps de réparation T_r; si la loi n'est pas exponentielle et qu'un élément soit en réparation à l'instant t_0, le temps que durera encore la réparation en cours dépendra de l'instant où elle a été commencée, par suite, ce processus ne sera non plus markovien.

En général, la loi exponentielle joue un rôle particulièrement important dans la théorie des processus markoviens à temps continu. Il est facile de voir que dans un processus markovien stationnaire le temps durant lequel le système reste dans un état quelconque est tou-

*) On suppose que les défauts ne peuvent survenir que quand le système fonctionne; la probabilité de deux ou plusieurs défauts simultanés est négligeable.

jours réparti suivant une loi exponentielle (dont le paramètre dépend en général de cet état). En effet, supposons qu'à un instant t_0 le système se trouve dans l'état x_k et ceci depuis un certain temps. Par définition d'un processus de Markov, la probabilité d'un événement quelconque dans le futur ne dépend pas de la préhistoire du processus ; en particulier, la probabilité pour que le système sorte de l'état x_k durant le temps t n'a aucune raison de dépendre du temps que le système a passé dans cet état. Par conséquent, le temps que le système passera dans l'état x_k doit suivre une loi exponentielle.

Lorsque le processus se déroulant dans un système physique à états dénombrables et à temps continu est markovien, il peut être décrit par des équations différentielles ordinaires où les fonctions inconnues sont les probabilités $p_1(t)$, $p_2(t)$, . . . des états. La manière d'écrire et de résoudre ces équations va être exposée dans le paragraphe suivant sur l'exemple d'un système d'attente simple.

19.8. Système d'attente à demandes refusées. Equations d'Erlang

Les systèmes d'attente peuvent être classés en deux groupes: a) les systèmes à demandes refusées et b) les systèmes à file d'attente.

Dans les systèmes du premier groupe une demande apparaissant à un instant où tous les canaux sont occupés est immédiatement refusée, elle quitte le système et n'est plus comptée.

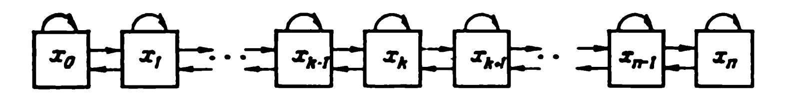


Fig. 19.8.1

Dans les systèmes du deuxième groupe une demande trouvant tous les canaux occupés ne quitte pas le système, mais prend sa place dans la file en attendant qu'un canal quelconque se soit disponible. Nous commençons par le cas le plus simple, le cas d'un système à demandes refusées.

Soit un système d'attente à n canaux à demandes refusées. Nous allons l'envisager comme un système physique X à un nombre fini d'états:

x_0 — tous les canaux sont libres;
x_1 — un seul canal est occupé;
. . . .
x_k — k canaux sont occupés;
. . . .
x_n — tous les n canaux sont occupés.

Le schéma des transitions possibles est donné sur la figure 19.8.1.

Proposons-nous de déterminer les probabilités $p_k(t)$ $(k = 0, 1, \ldots, n)$ des états du système pour un instant quelconque t. La solution est basée sur les hypothèses suivantes:

1) le flux de demandes est simple, de densité λ:

2) le temps de service T_{ser} suit une loi exponentielle de paramètre $\mu = \frac{1}{m_{t_{ser}}}$:

$$g(t) = \mu e^{-\mu t} \quad (t > 0). \tag{19.8.1}$$

Notons que le paramètre μ dans la formule (19.8.1) est analogue au paramètre λ de la loi exponentielle de répartition de l'intervalle T entre les événements voisins d'un flux simple:

$$f(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{19.8.2}$$

Le sens du paramètre λ est la « densité du flux de demandes ». D'une manière analogue la grandeur μ peut être interprétée comme la « densité du flux de disponibilités » d'un canal occupé. En effet, supposons

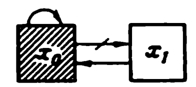


Fig. 19.8.2

qu'un canal soit occupé en permanence (les demandes arrivent d'une manière continue) ; il est clair qu'alors dans ce canal on aura un flux de « disponibilités » de densité μ.

Comme les deux flux sont simples, le processus se déroulant dans le système sera markovien.

Considérons les états possibles du système et leurs probabilités:

$$p_0(t), \quad p_1(t), \ \ldots, \ p_n(t). \tag{19.8.3}$$

Il est évident qu'à un instant quelconque on a:

$$\sum_{k=0}^{n} p_k(t) = 1. \tag{19.8.4}$$

Ecrivons les équations différentielles pour toutes les probabilités (19.8.3) en commençant par $p_0(t)$. Fixons l'instant t et recherchons la probabilité $p_0(t + \Delta t)$ pour qu'à l'instant $t + \Delta t$ le système se trouve dans l'état x_0 (tous les canaux sont libres). Ceci peut avoir lieu de deux façons différentes (fig. 19.8.2):

A — à l'instant t le système se trouvait dans l'état x_0 et durant le temps Δt il n'est pas passé à l'état x_1 (aucune demande n'a été reçue) ;

B — à l'instant t le système se trouvait dans l'état x_1, durant le temps Δt le canal s'est libéré et le système est passé à l'état x_0.

Sur la figure 19.8.2 on a montré par des flèches dirigées vers l'état x_0 comment le système peut y arriver à l'instant $t + \Delta t$; la flèche issue de x_0 est barrée pour marquer que le système ne doit pas quitter l'état x_0.

La probabilité de sauter un état (de passer, par exemple, directement de x_2 à x_0 sans s'arrêter dans x_1) durant un laps de temps court est négligeable devant les probabilités $P(A)$ et $P(B)$ *).

En vertu du théorème des probabilités totales on a:

$$p_0(t + \Delta t) \approx P(A) + P(B). \tag{19.8.5}$$

Nous allons utiliser le théorème des probabilités composées pour trouver la probabilité de l'événement A. La probabilité pour que, à l'instant t, le système se trouve dans l'état x_0 est égale à $p_0(t)$. La probabilité pour que, durant l'intervalle Δt, aucune demande ne soit reçue est égale à $e^{-\lambda\Delta t}$. On a approximativement:

$$e^{-\lambda\Delta t} \approx 1 - \lambda\Delta t. \tag{19.8.6}$$

Par conséquent,

$$P(A) \approx p_0(t)(1 - \lambda\Delta t).$$

Calculons $P(B)$. La probabilité pour qu'à l'instant t le système se trouve dans l'état x_1 est égale à $p_1(t)$. La probabilité pour que, durant l'intervalle Δt le canal soit disponible est égale à $1 - e^{-\mu\Delta t}$; on a approximativement:

$$1 - e^{-\mu\Delta t} \approx \mu\Delta t.$$

Par conséquent,

$$P(B) \approx p_1(t)\,\mu\Delta t.$$

D'où

$$p_0(t + \Delta t) \approx p_0(t)(1 - \lambda\,\Delta t) + \mu p_1(t)\,\Delta t.$$

En faisant passer $p_0(t)$ dans le premier membre, puis en divisant par Δt et en passant à la limite pour $\Delta t \to 0$, on obtient l'équation différentielle cherchée pour $p_0(t)$:

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t). \tag{19.8.7}$$

On peut trouver des équations différentielles analogues pour les probabilités des autres états.

Fixons un k $(0 < k < n)$ quelconque et calculons la probabilité $p_k(t + \Delta t)$ pour qu'à l'instant $t + \Delta t$ le système se trouve dans l'état x_k (fig. 19.8.3). Cette probabilité peut être calculée comme la

---

*) Dans la suite nous allons négliger les termes d'ordre de petitesse plus élevé que Δt. A la limite pour $\Delta t \to 0$ les égalités approchées deviendront exactes.

probabilité de la somme non plus de deux mais de trois événements (nombre de flèches dirigées vers l'état x_k):

A — à l'instant t le système se trouvait dans l'état x_k (k canaux occupés) et durant le temps Δt il n'est pas passé ni dans x_{k+1}, ni dans x_{k-1} (aucune demande n'est arrivée, aucun canal n'est devenu disponible);

B — à l'instant t le système se trouvait dans l'état x_{k-1} ($k-1$ canaux occupés) et durant le temps Δt il est passé dans x_k (une demande a été reçue);

C — à l'instant t le système se trouvait dans l'état x_{k+1} ($k+1$ canaux occupés) et durant le temps Δt un canal a été libéré.

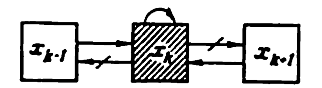


Fig. 19.8.3

Calculons $P(A)$. Recherchons tout d'abord la probabilité pour que durant le temps Δt aucune demande ne soit arrivée, aucun canal ne soit disponible:

$$e^{-\lambda\Delta t}(e^{-\mu\Delta t})^k = e^{-(\lambda+k\mu)\Delta t}.$$

En négligeant les infiniment petits d'ordre élevé on a:

$$e^{-(\lambda+k\mu)\Delta t} \approx 1-(\lambda+k\mu)\Delta t,$$

d'où

$$P(A) \approx p_k(t)[1-(\lambda+k\mu)\Delta t].$$

D'une manière analogue

$$P(B) \approx p_{k-1}(t)\lambda\Delta t,$$

$$P(C) \approx p_{k+1}(t)(k+1)\mu\Delta t$$

et

$$p_k(t+\Delta t) \approx p_k(t)[1-(\lambda+k\mu)\Delta t] + p_{k-1}(t)\lambda\Delta t + p_{k+1}(t)(k+1)\mu\Delta t.$$

D'où l'équation différentielle pour $p_k(t)$ $(0 < k < n)$:

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda+k\mu)p_k(t) + (k+1)\mu p_{k+1}(t).$$

Ecrivons l'équation correspondant à la dernière probabilité $p_n(t)$ (fig. 19.8.4). On a:

$$p_n(t+\Delta t) \approx p_n(t)(1-n\mu\Delta t) + p_{n-1}(t)\lambda\Delta t,$$

où $1-n\mu\Delta t$ est la probabilité pour que durant le temps Δt aucun canal ne soit disponible; $\lambda\Delta t$ la probabilité pour que durant ce temps Δt une demande arrive. On obtient ainsi l'équation différentielle

pour $p_n(t)$:

$$\frac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - n\mu p_n(t).$$

Ainsi, on a obtenu un système d'équations différentielles pour les probabilités $p_1(t)$, $p_2(t)$, . . ., $p_n(t)$:

$$\left.\begin{aligned}
&\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t),\\
&\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)\, p_k(t) +\\
&\quad + (k+1)\,\mu p_{k+1}(t) \quad (0 < k < n),\\
&\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - n\mu p_n(t).
\end{aligned}\right\} \qquad (19.8.8)$$

Les équations (19.8.8) sont appelées *équations d'Erlang*. En intégrant le système (19.8.8.) pour les conditions initiales

$$p_0(0) = 1; \quad p_1(0) = \dots = p_n(0) = 0$$

(à l'instant initial tous les canaux sont libres) on obtient $p_k(t)$ pour tout k. Les probabilités $p_k(t)$ caractérisent le taux moyen d'occupation du système et sa variation dans le temps. En particulier, $p_n(t)$

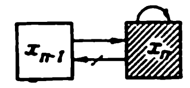


Fig. 19.8.4

est la probabilité pour que la demande arrivée à l'instant t trouve tous les canaux occupés (soit refusée):

$$P_{\text{ref}} = p_n(t).$$

La grandeur $q(t) = 1 - p_n(t)$ est appelée *pouvoir de transmission relatif* du système. A l'instant t c'est le rapport du nombre moyen de demandes servies par unité de temps au nombre moyen de demandes arrivées.

L'intégration du système d'équations différentielles linéaires (19.8.8) est assez facile pour un nombre quelconque de canaux n.

Notons que pour la déduction des équations (19.8.8) on n'a nulle part supposé que les grandeurs λ et μ (densité du flux de demandes et du flux de « disponibilités » sont constantes. Les équations (19.8.8) restent vraies lorsque $\lambda(t)$, $\mu(t)$ dépendent du temps, il suffit que les flux d'événements faisant passer le système d'un état à l'autre soient poissoniens (sans cela le processus ne sera pas markovien).

19.9. Régime permanent de service. Formules d'Erlang

Considérons un système d'attente à n canaux à demandes refusées dont l'entrée reçoit un flux de demandes simple de densité λ; le temps de service suit une loi exponentielle de paramètre μ. La question qui se pose est de savoir si le processus aléatoire se déroulant dans le système est stationnaire. Il est clair qu'au début, tout de suite après la mise en marche du système, le processus ne sera pas stationnaire : dans le système d'attente (comme dans tout système dynamique) apparaît un processus transitoire. Cependant quelque temps après ce processus s'amortit et régime de fonctionnement du système devient stationnaire ou permanent. Ses caractéristiques ne dépendent plus du temps.

Dans de nombreuses applications on a besoin de connaître les caractéristiques du régime de service permanent limite.

On peut montrer que dans tout système à demandes refusées ce régime limite existe, c'est-à-dire que pour $t \to \infty$ toutes les probabilités $p_0(t)$, $p_1(t)$, . . ., $p_n(t)$ tendent vers des limites constantes p_0, p_1, . . ., p_n et toutes leurs dérivées, vers zéro.

Pour trouver les probabilités limites p_0, p_1, . . ., p_n (probabilités des états du système en régime permanent) remplaçons dans les équations (19.8.8) toutes les probabilités $p_k(t)$ $(0 \leqslant k \leqslant n)$ par leurs limites p_k et supposons nulles toutes leurs dérivées. A la place d'un système d'équations différentielles on obtient un système d'équations algébriques, soit :

$$\left.\begin{array}{l}
-\lambda p_0 + \mu p_1 = 0, \\
\lambda p_0 - (\lambda + \mu) p_1 + 2\mu p_2 = 0, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\lambda p_{k-1} - (\lambda + k\mu) p_k + (k+1)\mu p_{k+1} = 0 \quad (0 < k < n), \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\lambda p_{n-2} - [\lambda + (n-1)\mu] p_{n-1} + n\mu p_n = 0, \\
\lambda p_{n-1} - n\mu p_n = 0.
\end{array}\right\} \qquad (19.9.1)$$

Il faut ajouter à ces équations la condition

$$\sum_{k=0}^{n} p_k = 1. \qquad (19.9.2)$$

Le système (19.9.1) est résolvable par rapport aux inconnues p_0, p_1, . . ., p_n. La première équation donne :

$$p_1 = \frac{\lambda}{\mu} p_0, \qquad (19.9.3)$$

la seconde, compte tenu de (19.9.3), donne:

$$p_2=\frac{1}{2\mu}\left[-\lambda p_0+(\lambda+\mu)\,p_1\right]=\frac{1}{2\mu}\left[-\lambda p_0+\frac{\lambda^2}{\mu}\,p_0+\lambda p_0\right]=\frac{\lambda^2}{2\mu^2}\,p_0; \tag{19.9.4}$$

d'une manière analogue, compte tenu de (19.9.3) et (19.9.4), on obtient à partir de la troisième équation:

$$p_3=\frac{1}{3\mu}\left(-\frac{\lambda^2}{\mu}\,p_0+\frac{\lambda^3}{2\mu^2}\,p_0+\frac{\lambda^2}{2\mu^2}\cdot 2\mu p_0\right)=\frac{\lambda^3}{3\,!\,\mu^3}\,p_0,$$

et pour $k\leqslant n$ quelconque

$$p_k=\frac{\lambda^k}{k\,!\,\mu^k}\,p_0. \tag{19.9.5}$$

Introduisons la désignation

$$\frac{\lambda}{\mu}=\alpha \tag{19.9.6}$$

et appelons la grandeur α *densité rapportée du flux de demandes.* Ce n'est rien d'autre que le nombre moyen de demandes rapporté au temps moyen de service d'une demande. En effet:

$$\alpha=\frac{\lambda}{\mu}=\lambda m_{t_{\text{ser}}},$$

où $m_{t_{\text{ser}}}=M\,[T_{\text{ser}}]$ est le nombre moyen de service d'une demande. Dans les nouvelles désignations la formule (19.9.5) devient:

$$p_k=\frac{\alpha^k}{k!}\,p_0. \tag{19.9.7}$$

La formule (19.9.7) donne toutes les probabilités p_k en fonction de p_0. Pour les exprimer directement en fonction de α et de n servons-nous de la condition (19.9.2). En y substituant (19.9.7) on obtient:

$$\sum_{k=0}^{n} p_k=p_0\sum_{k=0}^{n}\frac{\alpha^k}{k!}=1,$$

d'où

$$p_0=\frac{1}{\sum\limits_{k=0}^{n}\frac{\alpha^k}{k!}}. \tag{19.9.8}$$

En substituant (19.9.8) dans (19.9.7) on obtient finalement:

$$p_k=\frac{\frac{\alpha^k}{k!}}{\sum\limits_{k=0}^{n}\frac{\alpha^k}{k!}}\quad(0\leqslant k\leqslant n). \tag{19.9.9}$$

Les formules (19.9.9) sont appelées *formules d'Erlang*. Elles donnent la loi de répartition limite du nombre de canaux occupés en fonction des caractéristiques du flux de demandes et du taux de service du système. En posant dans la formule (19.9.9) $k = n$ on obtient la *probabilité d'un refus* (probabilité pour qu'une demande arrivée au système trouve tous les canaux occupés)

$$P_{ref} = p_n = \frac{\frac{\alpha^n}{n!}}{\sum_{k=0}^{n} \frac{\alpha^k}{k!}}. \tag{19.9.10}$$

En particulier, pour un système monocanal ($n = 1$) on a:

$$P_{ref} = p_1 = \frac{\alpha}{1+\alpha} \tag{19.9.11}$$

et le débit relatif

$$q = 1 - P_{ref} = \frac{1}{1+\alpha}. \tag{19.9.12}$$

Les formules d'Erlang (19.9.9) et les formules (19.9.10) à (19.9.12) en découlant ont été trouvées dans l'hypothèse que le temps de service suit une loi exponentielle. Cependant, il a été montré *) que ces formules restent vraies quelle que soit la loi de répartition du temps de service, à condition toutefois que le flux des arrivées soit simple.

Exemple. Une station téléphonique automatique a 4 lignes de communications. Les appels forment un flux simple de densité $\lambda = 3$ (appels par minute). Tout appel arrivé lorsque toutes les lignes sont occupées est refusé **). La durée moyenne de conversation est 2 minutes. Trouver: a) la probabilité d'un refus P_{ref} et b) la probabilité p_0 pour que la station téléphonique soit libre.

Solution. On a $m_{t_{ser}} = 2$ (mn);

$$\mu = 0{,}5 \text{ (conv/mn)}; \quad \alpha = \frac{\lambda}{\mu} = 6.$$

a) La formule (19.9.10) donne:

$$P_{ref} = p_4 = \frac{\frac{\alpha^4}{4!}}{1 + \frac{\alpha}{1} + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \frac{\alpha^4}{4!}} \approx 0{,}47.$$

b) La formule (19.9.8) donne:

$$p_0 = \frac{1}{1 + \frac{\alpha}{1} + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \frac{\alpha^4}{4!}} \approx 0{,}0087.$$

---

*) Voir, par exemple, B. Sévastianov. *Théorème ergodique pour les processus markoviens et son application aux systèmes téléphoniques à appels refusés* (en russe). « Théoria véroïatnostei i éye priménénia », vol. 2, n° 1, 1957.

**) On suppose que les appels répétés des abonnés refusés ne changent pas le caractère poissonien du flux de demandes.

19.10. Système d'attente à files

Un système d'attente est dit à files si aucune demande n'est refusée mais prend sa place dans la file en attendant d'être servie.

Si le temps d'attente d'une demande n'est pas limité, le système est appelé système à file illimitée. Si le temps d'attente est limité par une condition quelconque, le système est du type mixte. C'est le cas intermédiaire entre un système à demandes refusées et un système à file illimitée.

Dans les applications, les systèmes mixtes présentent le plus d'intérêt.

Les restrictions imposées à l'attente sont très variées. Souvent on limite le temps d'attente des demandes dans la file, en lui imposant une limite supérieure T_{at}, qui peut être tant déterministe qu'aléatoire. Cette limite concerne seulement le temps d'attente dans la file tandis que le service, une fois commencé, se termine indépendamment du temps passé à faire la queue (par exemple, un client ayant occupé un fauteuil chez un coiffeur ne le quitte plus jusqu'à la fin du service). Dans d'autres problèmes il est plus naturel de limiter non pas le temps d'attente mais le temps qu'une demande passe dans le système, y compris le service (par exemple, une cible aérienne ne peut se trouver dans la zone du tir qu'un temps limité et la quitte que le bombardement soit fini ou non). Enfin on peut envisager un système mixte (par exemple, les organisations vendant des produits qui ne sont pas de première nécessité) où une demande fait la queue seulement si celle-ci n'est pas trop longue. Ici la limitation est imposée au nombre de demandes dans la file.

Pour les systèmes d'attente à file, on introduit la notion très importante de « discipline d'attente ». Les demandes peuvent être servies dans l'ordre d'arrivée (celui qui est arrivé avant est servi avant) ou choisies au hasard. Il y a des systèmes où certaines demandes jouissent d'une priorité de service (par exemple, les malades accidentés sont admis d'urgence).

Chaque type de systèmes d'attente a ses particularités et sa théorie mathématique, la plupart d'entre eux sont étudiés dans le cours de B. Gnédenko « Leçons sur la théorie des phénomènes d'attente » (en russe), Kiev, 1960.

Nous n'envisagerons ici que le cas le plus simple d'un système mixte, qui est une généralisation naturelle du problème d'Erlang pour un système à demandes refusées. Nous écrirons pour ce cas des équations différentielles analogues aux équations d'Erlang, et les formules pour les probabilités des états dans le régime permanent analogues aux formules d'Erlang.

Considérons un système d'attente mixte X à n canaux pour les conditions suivantes. Le flux de demandes est simple de densité λ,

Le temps de service T_{ser} d'une demande suit une loi exponentielle de paramètre $\mu = \frac{1}{m_{t_{ser}}}$. Une demande ayant trouvé tous les canaux occupés attend d'être servie; le temps d'attente est limité à T_{at}, si avant ce temps la demande n'est pas servie, elle quitte la file. On suppose que le temps d'attente T_{at} est aléatoire et réparti suivant une loi exponentielle:

$$h(t) = \nu e^{-\nu t} \quad (t > 0),$$

où le paramètre ν est une grandeur inversement proportionnelle à la durée moyenne de l'attente:

$$\nu = \frac{1}{m_{t_{at}}}; \quad m_{t_{at}} = M[T_{at}].$$

Le paramètre ν peut être interprété comme la densité du flux de demandes quittant la file sans être servies, tout comme les paramètres λ et μ le sont respectivement pour le flux de demandes et le flux de « disponibilités » des canaux. En effet, imaginons une demande qui ne fait que prendre la place dans la file, attendre le temps T_{at} et partir pour revenir immédiatement dans la file. Alors le flux des départs d'une telle demande aura justement une densité égale à ν.

Il est évident que pour $\nu \to \infty$ un système du type mixte devient un système à demandes refusées; pour $\nu \to 0$ il devient un système à file illimitée.

Notons que lorsque la durée d'attente suit une loi exponentielle le taux de service du système ne dépend pas du fait que les demandes sont servies par ordre ou aléatoirement, car pour une demande quelconque la loi de répartition du temps qu'elle doit encore attendre dans la file ne dépend pas du temps que cette attente a déjà duré.

En supposant poissoniens tous les flux d'événements donnant lieu à des changements d'état du système, on obtient que tout processus se déroulant dans ce système est markovien. Nous allons écrire les équations pour les probabilités des états du système. A cet effet, énumérons tout d'abord ces états, en étendant la numération aux demandes dans la file. Nous dirons qu'une demande est « liée » au système si elle est en train d'être servie ou si elle attend son tour d'être servie. Les états possibles du système sont:

x_0 — aucun canal n'est occupé (aucune demande n'attend),
x_1 — un canal est occupé (aucune demande n'attend),
. . . .
x_k — k canaux sont occupés (aucune demande n'attend),
. . . .
x_n — tous les n canaux sont occupés (aucune demande n'attend),
x_{n+1} — tous les n canaux sont occupés (une demande attend),
. . . .

x_{n+s} — tous les n canaux sont occupés (s demandes attendent),

. . . .

Le nombre de demandes dans la file est illimité. Ainsi, le système X est à un nombre infini d'états (bien que dénombrable). Le nombre d'équations différentielles décrivant ce système sera également infini.

Il est clair que les n premières équations différentielles ne se distinguent en rien des équations correspondantes d'Erlang:

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t),$$

. . . .

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)\, p_k(t) + (k+1)\,\mu p_{k+1}(t),$$

. . . .

$$\frac{dp_{n-1}(t)}{dt} = \lambda p_{n-2}(t) - [\lambda + (n-1)\,\mu]\, p_{n-1}(t) + n\mu p_n(t).$$

La différence apparaîtra pour $k = n$. En effet, un système à demandes refusées peut arriver dans l'état x_n seulement à partir de l'état x_{n-1} ; au contraire, un système à file d'attente peut passer dans l'état x_n non seulement à partir de x_{n-1}, mais également à partir de x_{n+1} (tous les canaux sont occupés, une demande attend dans la file).

Nous allons écrire l'équation différentielle pour $p_n(t)$. Fixons un instant t et trouvons la probabilité $p_n(t + \Delta t)$ pour qu'à l'instant $t + \Delta t$ le système se trouve dans l'état x_n. Ceci peut se produire de trois manières différentes:

1) à l'instant t le système se trouvait déjà dans l'état x_n et durant Δt n'en est pas sorti (aucune demande n'est arrivée et aucun des canaux ne s'est libéré) ;

2) à l'instant t le système se trouvait dans l'état x_{n-1} et durant Δt est passé à l'état x_n (une demande est arrivée) ;

3) à l'instant t le système se trouvait dans l'état x_{n+1} (tous les canaux occupés, une demande attend) et durant Δt il est passé à l'état x_n (ou bien un canal s'est libéré et une demande de la file d'attente l'a occupé, ou bien une demande a quitté la file par suite de l'écoulement du temps d'attente).

On a

$$p_n(t+\Delta t) \approx p_n(t)\,(1 - \lambda\Delta t - n\mu\Delta t) + p_{n-1}(t)\,\lambda\Delta t + p_{n+1}(t)\,(n\mu + \nu)\,\Delta t,$$

d'où

$$\frac{dp_n(t)}{dt} = -(\lambda + n\mu)\, p_n(t) + \lambda p_{n-1}(t) + (n\mu + \nu)\, p_{n+1}(t).$$

Calculons maintenant $p_{n+s}(t + \Delta t)$ pour $s > 0$ quelconque, c'est-à-dire la probabilité pour que, à l'instant $t + \Delta t$ tous les n canaux soient occupés et s demandes soient dans la file d'attente. Ceci peut également avoir lieu de trois manières différentes:

1) à l'instant t le système se trouvait déjà dans l'état x_{n+s} et durant Δt cet état n'a pas changé (ce qui signifie qu'aucune demande n'est arrivée, ni aucune des s demandes dans la file ne l'a quittée);

2) à l'instant t le système se trouvait dans l'état x_{n+s-1}, et durant Δt il est passé à l'état x_{n+s} (une demande est arrivée);

3) à l'instant t le système se trouvait dans l'état x_{n+s+1} et durant Δt il est passé à l'état x_{n+s} (pour cela il faut qu'ou bien un canal se libère et alors une des $s+1$ demandes de la file l'occupe, ou bien une des demandes de la file parte sans être servie, le temps d'attente s'étant écoulé).

Par conséquent:

$$p_{n+s}(t+\Delta t) \approx p_{n+s}(t)(1-\lambda\Delta t-n\mu\Delta t-s\nu\Delta t)+p_{n+s-1}(t)\lambda\Delta t+ \\ +p_{n+s+1}(t)[n\mu+(s+1)\nu]\Delta t,$$

d'où

$$\frac{dp_{n+s}(t)}{dt}=-(\lambda+n\mu+s\nu)p_{n+s}(t)+ \\ +\lambda p_{n+s-1}(t)+[n\mu+(s+1)\nu]p_{n+s+1}(t).$$

On obtient ainsi pour les probabilités des états un système infini d'équations différentielles:

$$\left.\begin{aligned}
&\frac{dp_0(t)}{dt}=-\lambda p_0(t)+\mu p_1(t),\\
&\frac{dp_1(t)}{dt}=\lambda p_0(t)-(\lambda+\mu)p_1(t)+2\mu p_2(t),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_k(t)}{dt}=\lambda p_{k-1}(t)-(\lambda+k\mu)p_k(t)+(k+1)\mu p_{k+1}(t)\quad(1\leqslant k\leqslant n-1),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_n(t)}{dt}=\lambda p_{n-1}(t)-(\lambda+n\mu)p_n(t)+(n\mu+\nu)p_{n+1}(t),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_{n+s}(t)}{dt}=\lambda p_{n+s-1}(t)-(\lambda+n\mu+s\nu)p_{n+s}(t)+\\
&\qquad\qquad+[n\mu+(s+1)\nu]p_{n+s+1}(t),\\
&\dots\dots\dots\dots\dots\dots\dots\dots
\end{aligned}\right\}\tag{19.10.1}$$

Les équations (19.10.1) sont une généralisation naturelle des équations d'Erlang pour le cas d'un système mixte à temps d'attente limité. Les paramètres λ, μ, ν dans ces équations peuvent être tant constants que variables. Lors de l'intégration du système (19.10.1) il faut tenir compte du fait que, bien que le nombre théorique d'états possibles du système soit infini, en fait les probabilités $p_{n+s}(t)$, lorsque s augmente, deviennent infiniment petites et l'on peut ne pas tenir compte des équations correspondantes.

Ecrivons maintenant les formules analogues aux formules d'Erlang pour les probabilités des états du système en régime de service

(permanent pour $t \to \infty$). On peut à partir des équations (19.10.1), supposant toutes les p_k ($k = 0, \ldots, n, \ldots$) constantes et toutes les dérivées nulles, obtenir le système d'équations algébriques suivans:

$$\left.\begin{array}{c} -\lambda p_0 + \mu p_1 = 0, \\ \lambda p_0 - (\lambda + \mu) p_1 + 2\mu p_2 = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{k-1} - (\lambda + k\mu) p_k + (k+1) \mu p_{k+1} = 0 \\ (1 \leqslant k \leqslant n-1), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{n-1} - (\lambda + n\mu) p_n + (n\mu + \nu) p_{n+1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda p_{n+s-1} - (\lambda + n\mu + s\nu) p_{n+s} + [n\mu + (s+1)\nu] p_{n+s+1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{array}\right\} \quad (19.10.2)$$

A ces équations il faut ajouter la condition:

$$\sum_{k=0}^{\infty} p_k = 1. \quad (19.10.3)$$

Nous allons chercher la solution du système (19.10.2).

A cet effet nous allons procéder comme dans le cas d'un système à demandes refusées: nous allons résoudre la première équation par rapport à p_1, substituer la solution dans la seconde, etc. Pour $k \leqslant n$ quelconque, tout comme dans le cas d'un système à demandes refusées on a:

$$p_k = \frac{\lambda^k}{k!\mu^k} p_0. \quad (19.10.4)$$

D'une manière analogue, pour les équations pour $k > n$ ($k = n + s$, $s \geqslant 1$) on obtient:

$$p_{n+1} = \frac{\lambda^{n+1} p_0}{n!\mu^n (n\mu + \nu)},$$

$$p_{n+2} = \frac{\lambda^{n+2} p_0}{n!\mu^n (n\mu + \nu)(n\mu + 2\nu)},$$

et en général pour s quelconque $s \geqslant 1$:

$$p_{n+s} = \frac{\lambda^{n+s} p_0}{n!\mu^n \prod_{m=1}^{s} (n\mu + m\nu)} \quad (19.10.5)$$

La probabilité p_0 est présentée dans les deux formules (19.10.4) et (19.10.5). On peut la trouver à partir de la condition (19.10.3). En y substituant les expressions (19.10.4) et (19.10.5) pour $k \leqslant n$ et $s \geqslant 1$ on obtient

$$p_0 \left\{ \sum_{k=0}^{n} \frac{\lambda^k}{k!\mu^k} + \sum_{s=1}^{\infty} \frac{\lambda^{n+s}}{n!\mu^n \prod_{m=1}^{s} (n\mu + m\nu)} \right\} = 1,$$

d'où

$$p_0=\frac{1}{\sum_{k=0}^{n}\frac{\lambda^k}{k!\mu^k}+\sum_{s=1}^{\infty}\frac{\lambda^{n+s}}{n!\mu^n\prod_{m=1}^{s}(n\mu+m\nu)}}. \qquad (19.10.6)$$

Transformons les équations (19.10.4), (19.10.5) et (19.10.6) en y introduisant les densités rapportées α et β au lieu des densités λ et ν:

$$\left.\begin{aligned}\frac{\lambda}{\mu}&=\lambda m_{t_{ser}}=\alpha,\\ \frac{\nu}{\mu}&=\nu m_{t_{ser}}=\beta.\end{aligned}\right\} \qquad (19.10.7)$$

Les paramètres α et β expriment respectivement le nombre moyen d'arrivées et le nombre moyen de départs des demandes non servies par le temps moyen de service d'une demande.

Les équations (19.10.4), (19.10.5) et (19.10.6) s'écrivent alors:

$$p_k=\frac{\alpha^k}{k!}p_0 \quad (0<k\leqslant n); \qquad (19.10.8)$$

$$p_{n+s}=\frac{\frac{\alpha^{n+s}}{n!}p_0}{\prod_{m=1}^{s}(n+m\beta)}(s\geqslant 1); \qquad (19.10.9)$$

$$p_0=\frac{1}{\sum_{k=0}^{n}\frac{\alpha^k}{k!}+\frac{\alpha^n}{n!}\sum_{s=1}^{\infty}\frac{\alpha^s}{\prod_{m=1}^{s}(n+m\beta)}}. \qquad (19.10.10)$$

En substituant (19.10.10) dans (19.10.8) et (19.10.9) on obtient définitivement pour les probabilités des états du système les expressions:

$$p_k=\frac{\frac{\alpha^k}{k!}}{\sum_{k=0}^{n}\frac{\alpha^k}{k!}+\frac{\alpha^n}{n!}\sum_{s=1}^{\infty}\frac{\alpha^s}{\prod_{m=1}^{s}(n+m\beta)}} \quad (0\leqslant k\leqslant n); \qquad (19.10.11)$$

$$p_{n+s}=\frac{\frac{\alpha^n}{n!}\frac{\alpha^s}{\prod_{m=1}^{s}(n+m\beta)}}{\sum_{k=0}^{n}\frac{\alpha^k}{k!}+\frac{\alpha^n}{n!}\sum_{s=1}^{\infty}\frac{\alpha^s}{\prod_{m=1}^{s}(n+m\beta)}} \quad (s\geqslant 1). \qquad (19.10.12)$$

Connaissant les probabilités de tous les états du système on peut facilement trouver d'autres caractéristiques, en particulier la probabilité $P_{\text{dép}}$ pour qu'une demande quitte le système sans être servie. A cet effet nous raisonnons comme suit : en régime permanent la probabilité $P_{\text{dép}}$ n'est rien d'autre que le rapport du nombre moyen de demandes quittant la file d'attente par unité de temps au nombre moyen de demandes arrivées par unité de temps. Cherchons ce nombre moyen de départs par unité de temps. A cet effet calculons tout d'abord l'espérance mathématique m_s du nombre de demandes dans la file :

$$m_s = M[s] = \sum_{s=1}^{\infty} s p_{n+s} = \frac{\dfrac{\alpha^n}{n!} \displaystyle\sum_{s=1}^{\infty} \frac{s\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)}}{\displaystyle\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \sum_{s=1}^{\infty} \frac{\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)}}. \qquad (19.10.13)$$

Pour obtenir $P_{\text{dép}}$ il faut multiplier m_s par la « densité moyenne de départs » ν d'une demande et le diviser par la densité moyenne des arrivées λ, c'est-à-dire introduire le coefficient :

$$\frac{\nu}{\lambda} = \frac{\dfrac{\nu}{\mu}}{\dfrac{\lambda}{\mu}} = \frac{\beta}{\alpha}.$$

On a alors :

$$P_{\text{dép}} = \frac{\beta}{\alpha} \cdot \frac{\dfrac{\alpha^n}{n!} \displaystyle\sum_{s=1}^{\infty} \frac{s\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)}}{\displaystyle\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \sum_{s=1}^{\infty} \frac{\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)}}. \qquad (19.10.14)$$

Le débit relatif du système est caractérisé par la probabilité pour toute demande arrivée au système d'être servie :

$$q = 1 - P_{\text{dép}}.$$

Il est clair que le débit d'un système, à file d'attente pour des λ et μ donnés, est toujours supérieur au débit d'un système à demandes refusées : en effet, les demandes arrivées dans le système lorsque tous ses n canaux sont occupés ne quittent pas tout le système mais seulement une partie d'entre elles. Le débit augmente avec le temps moyen d'attente $m_{t_{\text{at}}} = \dfrac{1}{\nu}$.

Il est quelque peu difficile d'utiliser les formules (19.10.11), (19.10.12) et (19.10.14) car elles contiennent des sommes infinies. Cependant, les termes de ces sommes décroissent rapidement en valeur *).

Voyons ce que deviennent les formules (19.10.11) et (19.10.12) pour $\beta \to \infty$ et $\beta \to 0$. Il est évident que pour $\beta \to \infty$ un système à file d'attente devient un système à demandes refusées (une demande trouvant le système occupé le quitte immédiatement). En effet, pour $\beta \to \infty$ les formules (19.10.12) donneront des zéros, et les formules (19.10.11) deviendront les formules d'Erlang pour un système à demandes refusées.

Dans l'autre cas limite où $\beta \to 0$, les demandes arrivées ne quittent jamais la file, donc $P_{\text{dép}} = 0$; tôt ou tard toute demande est servie. Par contre, dans un système à file illimitée le régime permanent pour $t \to \infty$ n'existe pas toujours. On peut montrer qu'un tel régime existe seulement pour $\alpha < n$, c'est-à-dire lorsque le nombre moyen d'arrivées rapporté au temps de service d'une demande, n'est pas au-delà des limites des possibilités d'un système à n canaux. Pour $\alpha \geqslant n$, le nombre de demandes dans la file augmentera indéfiniment avec le temps.

Supposons que $\alpha < n$ et calculons les probabilités limites p_k $(0 \leqslant k \leqslant n)$ dans le cas d'un système à file illimitée. A cet effet dans les formules (19.9.10), (19.9.11) et (19.9.12) posons $\beta \to 0$. Il vient:

$$p_0 = \frac{1}{\sum\limits_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \sum\limits_{s=1}^{\infty} \frac{\alpha^s}{n^s}},$$

ou en prenant la somme de la progression (ce qui n'est possible que pour $\alpha < n$):

$$p_0 = \frac{1}{\sum\limits_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^{n+1}}{n!\,(n-\alpha)}} \tag{19.10.15}$$

*) Une estimation grossière de l'erreur due au fait de négliger tous les termes de la somme à partir du r-ième peut être donnée par les formules suivantes:

$$\sum_{s=r}^{\infty} \frac{\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)} < \frac{\left(\frac{\alpha}{\beta}\right)^r}{r!} e^{\frac{\alpha}{\beta}},$$

$$\sum_{s=r}^{\infty} \frac{s\alpha^s}{\prod\limits_{m=1}^{s} (n+m\beta)} < \frac{\left(\frac{\alpha}{\beta}\right)^r}{(r-1)!} e^{\frac{\alpha}{\beta}}.$$

D'où, en utilisant les formules (19.10.8) et (19.10.9), on obtient :

$$p_k = \frac{\dfrac{\alpha^k}{k!}}{\sum\limits_{k=1}^{n} \dfrac{\alpha^k}{k!} + \dfrac{\alpha^{n+1}}{n!\,(n-\alpha)}} \quad (0 \leqslant k \leqslant n), \tag{19.10.16}$$

et d'une manière analogue pour $k = n + s$ $(s > 0)$

$$p_{n+s} = \frac{\dfrac{\alpha^{n+s}}{n!\,n^s}}{\sum\limits_{k=0}^{n} \dfrac{\alpha^k}{k!} + \dfrac{\alpha^{n+1}}{n!\,(n-\alpha)}}. \tag{19.10.17}$$

Le nombre moyen de demandes dans la file est donné par la formule (19.10.13) pour $\beta \to 0$:

$$m_s = \frac{\dfrac{\alpha^{n+1}}{n \cdot n! \left(1 - \dfrac{\alpha}{n}\right)^2}}{\sum\limits_{k=0}^{n} \dfrac{\alpha^k}{k!} + \dfrac{\alpha^{n+1}}{n!\,(n-\alpha)}}. \tag{19.10.18}$$

E x e m p l e. Dans un système à trois canaux à temps d'attente illimité arrive un flux de demandes simple de densité $\lambda = 4$ (demandes par heure). Le temps moyen de service d'une demande est $m_{t_{ser}} = 30$ mn. Trouver si le régime permanent de service est possible ; si oui, trouver les probabilités p_0, p_1, p_2, p_3, la probabilité d'une file et sa longueur moyenne m_s.

S o l u t i o n. On a $\mu = \dfrac{1}{m_{t_{ser}}} = 2$, $\alpha = \dfrac{\lambda}{\mu} = 2$. Comme $\alpha < n$, le régime permanent est possible. La formule (19.10.16) permet de trouver :

$$p_0 = \frac{1}{9} \approx 0{,}111; \qquad p_1 = \frac{2}{9} \approx 0{,}222;$$

$$p_2 = \frac{2}{9} \approx 0{,}222; \qquad p_3 = \frac{8}{54} \approx 0{,}148.$$

La probabilité d'une file est :

$$P_{\text{file}} = 1 - (p_0 + p_1 + p_2 + p_3) = 0{,}297.$$

La longueur moyenne de la file d'attente est donnée par la formule (19.10.18) :

$$m_s \approx 0{,}89 \text{ (demande).}$$

19.11. Système du type mixte à longueur de file limitée

Dans le paragraphe précédent nous avons étudié un système à temps d'attente limité. Ici nous allons envisager un système du type mixte où la limitation concerne le nombre d'unités dans la file

d'attente. Supposons qu'une demande ayant trouvé tous les canaux occupés ne fasse la queue que si celle-ci contient moins de m demandes; si au contraire ce nombre est égal à m (il ne peut pas lui être supérieur) la dernière arrivée quitte le système sans être servie. Les autres hypothèses, celles sur les arrivées poissoniennes et la répartition exponentielle du temps de service, restent inchangées.

Ainsi, on a un système à n canaux à file d'attente, dans lequel le nombre de demandes dans la file est limité et égal à m. Nous allons écrire les équations différentielles pour les probabilités d'états du système. Notons que dans le cas considéré le nombre d'états du système sera fini, car le nombre total de demandes acceptées par le système ne peut être supérieur à $n + m$ (n en cours de service et m dans la file). Les états possibles du système sont:

x_0 — tous les canaux sont libres, aucune demande n'attend,
x_1 — un canal est occupé, » » »
. . . .
x_k — k canaux sont occupés, » » »
. . . .
x_{n-1} — $n - 1$ canaux sont occupés, » » »
x_n — tous les n canaux sont occupés, » » »
x_{n+1} — tous les n canaux sont occupés, une demande dans la file,
. . . .
x_{n+m} — tous les n canaux sont occupés, m demandes dans la file.

Il est évident que les n premières équations pour les probabilités $p_0(t), \ldots, p_{n-1}(t)$ coïncident avec les équations d'Erlang (19.8.8). Nous allons trouver les autres équations. On a:

$$p_n(t+\Delta t) = p_{n-1}(t)\lambda\Delta t + p_n(t)(1-\lambda\Delta t - n\mu\Delta t) + p_{n+1}(t)\, n\mu\Delta t,$$

d'où

$$\frac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - (\lambda + n\mu)\, p_n(t) + n\mu p_{n+1}(t).$$

Trouvons maintenant l'équation pour $p_{n+s}(t)$ $(1 \leqslant s < m)$:

$$p_{n+s}(t+\Delta t) = \\ = p_{n+s-1}(t)\lambda\Delta t + p_{n+s}(t)(1-\lambda\Delta t - n\mu\Delta t) + p_{n+s+1}(t)\, n\mu\Delta t,$$

d'où

$$\frac{dp_{n+s}(t)}{dt} = \lambda p_{n+s-1}(t) - (\lambda + n\mu)\, p_{n+s}(t) + n\mu p_{n+s+1}(t).$$

La dernière équation sera:

$$\frac{dp_{n+m}(t)}{dt} = \lambda p_{n+m-1}(t) - n\mu p_{n+m}(t).$$

On a ainsi obtenu un système de $(n + m + 1)$ équations différentielles :

$$\left.\begin{aligned}
&\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)\, p_k(t) + (k+1)\,\mu p_{k+1}(t) \quad (0 < k \leqslant n-1),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - (\lambda + n\mu)\, p_n(t) + n\mu p_{n+1}(t),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_{n+s}(t)}{dt} = \lambda p_{n+s-1}(t) - (\lambda + n\mu)\, p_{n+s}(t) + n\mu p_{n+s+1}(t) \quad (1 \leqslant s < m),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\frac{dp_{n+m}(t)}{dt} = \lambda p_{n+m-1} - n\mu p_{n+m}(t).
\end{aligned}\right\} \tag{19.11.1}$$

Considérons le cas limite pour $t \to \infty$. En annulant toutes les dérivées et supposant toutes les probabilités constantes, on obtient le système d'équations algébriques suivant :

$$\left.\begin{aligned}
&-\lambda p_0 + \mu p_1 = 0,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\lambda p_{k-1} - (\lambda + k\mu)\, p_k + (k+1)\,\mu p_{k+1} = 0 \quad (0 < k \leqslant n-1),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\lambda p_{n-1} - (\lambda + n\mu)\, p_n + n\mu p_{n+1} = 0,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\lambda p_{n+s-1} - (\lambda + n\mu)\, p_{n+s} + n\mu p_{n+s+1} = 0 \quad (1 \leqslant s < m),\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&\lambda p_{n+m-1} - n\mu p_{n+m} = 0
\end{aligned}\right\} \tag{19.11.2}$$

et la condition supplémentaire :

$$\sum_{k=0}^{n+m} p_k = 1. \tag{19.11.3}$$

Les équations (19.11.2) se résolvent tout comme les équations algébriques analogues des paragraphes précédents ; sans nous arrêter sur les opérations intermédiaires, nous donnons ici les formules défi-

nitives, soit:

$$p_k = \frac{\frac{\alpha^k}{k!}}{\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \sum_{s=1}^{m} \left(\frac{\alpha}{n}\right)^s} \quad (0 \leqslant k \leqslant n), \qquad (19.11.4)$$

$$p_{n+s} = \frac{\frac{\alpha^n}{n!} \left(\frac{\alpha}{n}\right)^s}{\sum_{k=0}^{n} \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \sum_{s=1}^{m} \left(\frac{\alpha}{n}\right)^s} \quad (1 \leqslant s \leqslant m). \qquad (19.11.5)$$

La probabilité pour qu'une demande quitte le système sans être servie est égale à la probabilité p_{n+m} que la file contienne déjà m demandes.

Il est facile de voir que les formules (19.11.4) et (19.11.5) s'obtiennent à partir des formules (19.10.11) et (19.10.12) en y posant $\beta = 0$ et en limitant la sommation à $s = m$.

Exemple. Une station de service d'automobiles reçoit un flux de demandes simple de densité $\lambda = 0{,}5$ (voitures par heure). La station dispose d'un seul atelier de réparation; de plus, trois voitures aux maximum peuvent attendre dans la cour leur tour d'être servies. Le temps moyen de réparation d'une voiture est $m_{t_{\text{ser}}} = \frac{1}{\mu} = 2$ (heures). Trouver: a) le débit du système; b) le temps moyen où la station est inoccupée; c) voir comment changeront ces caractéristiques si on ajoute encore un atelier de réparation.

Solution. On a: $\lambda = 0{,}5$; $\mu = 0{,}5$; $\alpha = 1$; $m = 3$.

a) En posant $n = 1$ dans la formule (19.11.5) on trouve la probabilité pour la voiture de quitter le système sans être servie:

$$P_{\text{dép}} = p_{1+3} = \frac{1}{1+1+3} = 0{,}20.$$

Le débit relatif du système est $q = 1 - P_{\text{dép}} = 0{,}80$. Le débit absolu: $Q = \lambda q = 0{,}4$ (voitures par heure).

b) Le temps moyen où la station est inoccupée est donné par la formule (19.11.4):

$$p_0 = \frac{1}{5} = 0{,}20.$$

c) En posant $n = 2$ on trouve:

$$P_{\text{dép}} = p_{2+3} = \frac{\frac{1}{16}}{1+1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}} = \frac{1}{47} \approx 0{,}021,$$

$q = 1 - P_{\text{dép}} \approx 0{,}979$ (c'est-à-dire qu'environ 98 % de toutes les demandes de service seront satisfaites):

$$Q = \lambda q \approx 0{,}49 \text{ (voitures par heure).}$$

Le temps moyen d'inoccupation est $p_0 = \frac{16}{47} \approx 0{,}34$, c'est-à-dire qu'à peu près 34 % du temps les ateliers seront inoccupés.

ANNEXES

Table 1

Valeurs de la fonction de répartition normale

$$\Phi^*(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt$$

x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ
—0,00	0,5000	40	—0,30	0,3821	38	—0,60	0,2743	34	—0,90	0,1841	27
—0,01	4960	40	—0,31	3783	38	—0,61	2709	33	—0,91	1814	26
—0,02	4920	40	—0,32	3745	38	—0,62	2676	33	—0,92	1788	26
—0,03	4880	40	—0,33	3707	38	—0,63	2643	32	—0,93	1762	26
—0,04	4840	39	—0,34	3669	37	—0,64	2611	33	—0,94	1736	25
—0,05	4801	40	—0,35	3632	38	—0,65	2578	32	—0,95	1711	26
—0,06	4761	40	—0,36	3594	37	—0,66	2546	32	—0,96	1685	25
—0,07	4721	40	—0,37	3557	37	—0,67	2514	31	—0,97	1660	25
—0,08	4681	40	—0,38	3520	37	—0,68	2483	32	—0,98	1635	24
—0,09	4641	39	—0,39	3483	37	—0,69	2451	31	—0,99	1611	24
—0,10	0,4602	40	—0,40	0,3446	37	—0,70	0,2420	31	—1,00	0,1587	24
—0,11	4562	40	—0,41	3409	37	—0,71	2389	31	—1,01	1563	24
—0,12	4522	39	—0,42	3372	36	—0,72	2358	31	—1,02	1539	24
—0,13	4483	40	—0,43	3336	36	—0,73	2327	30	—1,03	1515	23
—0,14	4443	39	—0,44	3300	36	—0,74	2297	31	—1,04	1492	23
—0,15	4404	40	—0,45	3264	36	—0,75	2266	30	—1,05	1469	23
—0,16	4364	39	—0,46	3228	36	—0,76	2236	30	—1,06	1446	23
—0,17	4325	39	—0,47	3192	36	—0,77	2206	29	—1,07	1423	22
—0,18	4286	39	—0,48	3156	35	—0,78	2177	29	—1,08	1401	22
—0,19	4247	40	—0,49	3121	36	—0,79	2148	29	—1,09	1379	22
—0,20	0,4207	39	—0,50	0,3085	35	—0,80	0,2119	29	—1,10	0,1357	22
—0,21	4168	39	—0,51	3050	35	—0,81	2090	29	—1,11	1335	21
—0,22	4129	39	—0,52	3015	34	—0,82	2061	28	—1,12	1314	22
—0,23	4090	38	—0,53	2981	35	—0,83	2033	28	—1,13	1292	21
—0,24	4052	39	—0,54	2946	34	—0,84	2005	28	—1,14	1271	20
—0,25	4013	39	—0,55	2912	35	—0,85	1977	28	—1,15	1251	21
—0,26	3974	38	—0,56	2877	34	—0,86	1949	27	—1,16	1230	20
—0,27	3936	39	—0,57	2843	33	—0,87	1922	28	—1,17	1210	20
—0,28	3897	38	—0,58	2810	34	—0,88	1894	27	—1,18	1190	20
—0,29	3859	38	—0,59	2776	33	—0,89	1867	26	—1,19	1170	19

Suite table 1

x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ	x	$\Phi^*(x)$	Δ
—1,20	0,1151	20	—1,65	0,0495	10	—3,00	0,0014	4	0,35	0,6368	38
—1,21	1131	19	—1,66	0485	10	—3,10	0010	3	0,36	6406	37
—1,22	1112	19	—1,67	0475	10	—3,20	0007	2	0,37	6443	37
—1,23	1093	18	—1,68	0465	10	—3,30	0005	2	0,38	6480	37
—1,24	1075	19	—1,69	0455	9	—3,40	0003	1	0,39	6517	37
—1,25	1056	18	—1,70	0,0446	10	—3,50	0002	0	0,40	0,6554	37
—1,26	1038	18	—1,71	0436	9	—3,60	0002	1	0,41	6591	37
—1,27	1020	17	—1,72	0427	9	—3,70	0001	0	0,42	6628	36
—1,28	1003	18	—1,73	0418	9	—3,80	0001	1	0,43	6664	36
—1,29	0985	17	—1,74	0409	8	—3,90	0000		0,44	6700	36
—1,30	0,0968	17	—1,75	0401	9	0,00	0,5000	40	0,45	6736	36
—1,31	0951	17	—1,76	0392	8	0,01	5040	40	0,46	6772	36
—1,32	0934	16	—1,77	0384	9	0,02	5080	40	0,47	6808	36
—1,33	0918	17	—1,78	0375	8	0,03	5120	40	0,48	6844	35
—1,34	0901	16	—1,79	0367	8	0,04	5160	39	0,49	6879	36
—1,35	0885	16	—1,80	0,0359	8	0,05	5199	40	0,50	0,6915	35
—1,36	0869	16	—1,81	0351	7	0,06	5239	40	0,51	6950	35
—1,37	0853	15	—1,82	0344	8	0,07	5279	40	0,52	6985	34
—1,38	0838	15	—1,83	0336	7	0,08	5319	40	0,53	7019	35
—1,39	0823	15	—1,84	0329	7	0,09	5359	39	0,54	7054	34
—1,40	0,0808	15	—1,85	0322	8	0,10	0,5398	40	0,55	7088	35
—1,41	0793	15	—1,86	0314	7	0,11	5438	40	0,56	7123	34
—1,42	0778	14	—1,87	0307	6	0,12	5478	39	0,57	7157	33
—1,43	0764	15	—1,88	0301	7	0,13	5517	40	0,58	7190	34
—1,44	0749	14	—1,89	0294	6	0,14	5557	39	0,59	7224	33
—1,45	0735	14	—1,90	0,0288	7	0,15	5596	40	0,60	0,7257	34
—1,46	0721	13	—1,91	0281	7	0,16	5636	39	0,61	7291	33
—1,47	0708	14	—1,92	0274	6	0,17	5675	39	0,62	7324	33
—1,48	0694	13	—1,93	0268	6	0,18	5714	39	0,63	7357	32
—1,49	0681	13	—1,94	0262	6	0,19	5753	40	0,64	7389	33
—1,50	0,0668	13	—1,95	0256	6	0,20	0,5793	39	0,65	7422	32
—1,51	0655	12	—1,96	0250	6	0,21	5832	39	0,66	7454	32
—1,52	0643	13	—1,97	0244	5	0,22	5871	39	0,67	7486	31
—1,53	0630	12	—1,98	0239	6	0,23	5910	38	0,68	7517	32
—1,54	0618	12	—1,99	0233	5	0,24	5948	39	0,69	7549	31
—1,55	0606	12	—2,00	0,0228	49	0,25	5987	39	0,70	0,7580	31
—1,56	0594	12	—2,10	0179	40	0,26	6026	38	0,71	7611	31
—1,57	0582	11	—2,20	0139	32	0,27	6064	39	0,72	7642	31
—1,58	0571	12	—2,30	0107	25	0,28	6103	38	0,73	7673	30
—1,59	0559	11	—2,40	0082	20	0,29	6141	38	0,74	7703	31
—1,60	0,0548	11	—2,50	0062	15	0,30	0,6179	38	0,75	7734	30
—1,61	0537	11	—2,60	0047	12	0,31	6217	38	0,76	7764	30
—1,62	0526	10	—2,70	0035	9	0,32	6255	38	0,77	7794	29
—1,63	0516	11	—2,80	0026	7	0,33	6293	38	0,78	7823	29
—1,64	0505	10	—2,90	0019	5	0,34	6331	37	0,79	7852	29

Suite table 1

x	Φ* (x)	Δ	x	Φ* (x)	Δ	x	Φ* (x)	Δ	x	Φ* (x)	Δ
0,80	0,7881	29	1,15	0,8749	21	1,50	0,9332	13	1,85	0,9678	8
0,81	7910	29	1,16	8770	20	1,51	9345	12	1,86	9686	7
0,82	7939	28	1,17	8790	20	1,52	9357	13	1,87	9693	6
0,83	7967	28	1,18	8810	20	1,53	9370	12	1,88	9699	7
0,84	7995	28	1,19	8830	19	1,54	9382	12	1,89	9706	7
0,85	8023	28	1,20	0,8849	20	1,55	9394	12	1,90	0,9713	6
0,86	8051	27	1,21	8869	19	1,56	9406	12	1,91	9719	7
0,87	8078	28	1,22	8888	19	1,57	9418	11	1,92	9726	6
0,88	8106	27	1,23	8907	18	1,58	9429	12	1,93	9732	6
0,89	8133	26	1,24	8925	19	1,59	9441	11	1,94	9738	6
0,90	0,8159	27	1,25	8944	18	1,60	0,9452	11	1,95	9744	6
0,91	8186	26	1,26	8962	18	1,61	9463	11	1,96	9750	6
0,92	8212	26	1,27	8980	17	1,62	9474	10	1,97	9756	5
0,93	8238	26	1,28	8997	18	1,63	9484	11	1,98	9761	6
0,94	8264	25	1,29	9015	17	1,64	9495	10	1,99	9767	5
0,95	8289	26	1,30	0,9032	17	1,65	9505	10	2,00	0,9772	49
0,96	8315	25	1,31	9049	17	1,66	9515	10	2,10	9821	40
0,97	8340	25	1,32	9066	16	1,67	9525	10	2,20	9861	32
0,98	8365	24	1,33	9082	17	1,68	9535	10	2,30	9893	25
0,99	8389	24	1,34	9099	16	1,69	9545	9	2,40	9918	20
1,00	0,8413	24	1,35	9115	16	1,70	0,9554	10	2,50	9938	15
1,01	8437	24	1,36	9131	16	1,71	9564	9	2,60	9953	12
1,02	8461	24	1,37	9147	15	1,72	9573	9	2,70	9965	9
1,03	8485	23	1,38	9162	15	1,73	9582	9	2,80	9974	7
1,04	8508	23	1,39	9177	15	1,74	9591	8	2,90	9981	5
1,05	8531	23	1,40	0,9192	15	1,75	9599	9	3,00	0,9986	4
1,06	8554	23	1,41	9207	15	1,76	9608	8	3,10	9990	3
1,07	8577	22	1,42	9222	14	1,77	9616	9	3,20	9993	2
1,08	8599	22	1,43	9236	15	1,78	9625	8	3,30	9995	2
1,09	8621	22	1,44	9251	14	1,79	9633	8	3,40	9997	1
1,10	0,8643	22	1,45	9265	14	1,80	0,9641	8	3,50	9998	0
1,11	8665	21	1,46	9279	13	1,81	9649	7	3,60	9998	1
1,12	8686	22	1,47	9292	14	1,82	9656	8	3,70	9999	0
1,13	8708	21	1,48	9306	13	1,83	9664	7	3,80	9999	1
1,14	8729	20	1,49	9319	13	1,84	9671	7	3,90	1,0000	

Table 2

Valeurs de la fonction e^x

x	e^{-x}	Δ	x	e^{-x}	Δ	x	e^{-x}	Δ	x	e^{-x}	Δ
0,00	1,000	10	0,40	0,670	7	0,80	0,449	4	3,00	0,050	5
0,01	0,990	10	0,41	0,664	7	0,81	0,445	5	3,10	0,045	4
02	980	10	42	657	7	82	440	4	3,20	41	4
03	970	9	43	650	6	83	436	4	3,30	37	4
04	961	10	44	644	6	84	432	5	3,40	33	3
05	951	9	45	638	7	85	427	4	3,50	30	3
06	942	10	46	631	6	86	423	4	3,60	27	2
07	932	9	47	625	6	87	419	4	3,70	25	3
08	923	9	48	619	6	88	415	4	3,80	22	2
09	914	9	49	613	7	89	411	4	3,90	20	2
0,10	0,905	9	0,50	0,606	6	0,90	0,407	4	4,00	0,0183	17
11	896	9	51	600	5	91	403	4	4,10	166	16
12	887	9	52	595	6	92	399	4	4,20	150	14
13	878	9	53	589	6	93	395	4	4,30	136	13
14	869	8	54	583	6	94	391	4	4,40	123	12
15	861	9	55	577	6	95	387	4	4,50	111	10
16	852	8	56	571	6	96	383	4	4,60	101	10
17	844	9	57	565	5	97	379	4	4,70	0,0091	9
18	835	8	58	560	6	98	375	3	4,80	82	8
19	827	8	59	554	5	99	372	4	4,90	74	7
0,20	0,819	8	0,60	0,549	6	1,00	0,368	35	5,00	0,0067	6
21	811	8	61	543	5	1,10	333	31	5,10	61	6
22	803	8	62	538	5	1,20	302	29	5,20	55	5
23	795	8	63	533	6	1,30	273	26	5,30	50	5
24	787	8	64	527	5	1,40	247	24	5,40	45	4
25	779	8	65	522	5	1,50	223	21	5,50	41	4
26	771	8	66	517	5	1,60	202	19	5,60	37	4
27	763	7	67	512	5	1,70	183	18	5,70	33	3
28	756	8	68	507	5	1,80	165	15	5,80	30	3
29	748	7	69	502	5	1,90	150	15	5,90	27	2
0,30	0,741	8	0,70	0,497	5	2,00	0,135	13	6,00	0,0025	3
31	733	7	71	492	5	2,10	122	11	6,10	22	2
32	726	7	72	487	5	2,20	111	11	6,20	20	2
33	719	7	73	482	5	2,30	100	9	6,30	18	1
34	712	7	74	477	5	2,40	0,091	9	6,40	17	2
35	705	7	75	472	4	2,50	82	8	6,50	15	1
36	698	7	76	468	5	2,60	74	7	6,60	14	2
37	691	7	77	463	5	2,70	67	6	6,70	12	1
38	684	7	78	458	4	2,80	61	6	6,80	11	1
39	677	7	79	454	5	2,90	55	5	6,90	10	1
0,40	0,670		0,80	0,449		3,00	0,050		7,00	0,0009	

Valeurs de la fonction $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9	x
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973	0,0
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918	0,1
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825	0,2
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697	0,3
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538	0,4
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352	0,5
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144	0,6
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920	0,7
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685	0,8
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444	0,9
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203	1,0
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965	1,1
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736	1,2
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518	1,3
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315	1,4
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127	1,5
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957	1,6
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804	1,7
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669	1,8
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551	1,9
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449	2,0
2,1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363	2,1
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290	2,2
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229	2,3
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180	2,4
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139	2,5
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107	2,6
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081	2,7
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061	2,8
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046	2,9
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034	3,0
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025	3,1
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018	3,2
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013	3,3
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009	3,4
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006	3,5
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004	3,6
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003	3,7
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002	3,8
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001	3,9
x	0	1	2	3	4	5	6	7	8	9	x

Valeurs du χ^2 pour

r \ p	0,99	0,98	0,95	0,90	0,80	0,70	0,50
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35
7	1,239	1,564	2,17	2,83	3,82	4,67	6,35
8	1,646	2,03	2,73	3.49	4,59	5,53	7,34
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34
13	4,11	4,76	5,89	7,04	8,63	9,93	12,34
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34
17	6,41	7,26	8,67	10,08	12,00	13,53	16,34
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34
19	7,63	8,57	10,11	11,65	13,72	15,35	18,34
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34
21	8,90	9,92	11,59	13,24	15,44	17,18	20,3
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3
23	10,20	11,29	13,09	14,85	17,19	19,02	22,3
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3
25	11,52	12,70	14,61	16,47	18,94	20,9	24,3
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3
27	12,88	14,12	16,15	18,11	20,7	22,7	26,3
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3
29	14,26	15,57	17,11	19,77	22,5	24,6	28,3
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3

Table 4

différents *r* et *p*

0,30	0,20	0,10	0,05	0,02	0,01	0,001	p / r
1,074	1,642	2,71	3,84	5,41	6,64	10,83	1
2,41	3,22	4,60	5,99	7,82	9,21	13,82	2
3,66	4,64	6,25	7,82	9,84	11,34	16,27	3
4,88	5,99	7,78	9,49	11,67	13,28	18,46	4
6,06	7,29	9,24	11,07	13,39	15,09	20,5	5
7,23	8,56	10,64	12,59	15,03	16,81	22,5	6
8,38	9,80	12,02	14,07	16,62	18,48	24,3	7
9,52	11,03	13,36	15,51	18,17	20,1	26,1	8
10,66	12,24	14,68	16,92	19,68	21,7	27,9	9
11,78	13,44	15,99	18,31	21,2	23,2	29,6	10
12,90	14,63	17,28	19,68	22,6	24,7	31,3	11
14,01	15,81	18,55	21,0	24,1	26,2	32,9	12
15,12	16,98	19,81	22,4	25,5	27,7	34,6	13
16,22	18,15	21,1	23,7	26,9	29,1	36,1	14
17,32	19,31	22,3	25,0	28,3	30,6	37,7	15
18,42	20,5	23,5	26,3	29,6	32,0	39,3	16
19,51	21,6	24,8	27,6	31,0	33,4	40,8	17
20,6	22,8	26,0	28,9	32,3	34,8	42,3	18
21,7	23,9	27,2	30,1	33,7	36,2	43,8	19
22,8	25,0	28,4	31,4	35,0	37,6	45,3	20
23,9	26,2	29,6	32,7	36,3	38,9	46,8	21
24,9	27,3	30,8	33,9	37,7	40,3	48,3	22
26,0	28,4	32,0	35,2	39,0	41,6	49,7	23
27,1	29,6	33,2	36,4	40,3	43,0	51,2	24
28,2	30,7	34,4	37,7	41,7	44,3	52,6	25
29,2	31,8	35,6	38,9	42,9	45,6	54,1	26
30,3	32,9	36,7	40,1	44,1	47,0	55,5	27
31,4	34,0	37,9	41,3	45,4	48,3	56,9	28
32,5	35,1	39,1	42,6	46,7	49,6	58,3	29
33,5	36,2	40,3	43,8	48,0	50,9	59,7	30

Valeurs de t_β satisfaisant à l'égalité $2\int_0^{t_\beta} S_{n-1}(t)\,dt = \beta$

$n-1$ \ β	0,1	0,2	0,3	0,4	0,5	0,6	0,7
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963
2	142	289	445	617	0,816	1,061	1,336
3	137	277	424	584	765	0,978	1,250
4	134	271	414	569	741	941	1,190
5	132	267	408	559	727	920	1,156
6	131	265	404	553	718	906	1,134
7	130	263	402	549	711	896	1,119
8	130	262	399	546	706	889	1,108
9	129	261	398	543	703	883	1,100
10	129	260	397	542	700	879	1,093
11	129	260	396	540	697	876	1,088
12	128	259	395	539	695	873	1,083
13	128	259	394	538	694	870	1,079
14	128	258	393	537	692	868	1,076
15	128	258	393	536	691	866	1,074
16	128	258	392	535	690	865	1,071
17	128	257	392	534	689	863	1,069
18	127	257	392	534	688	862	1,067
19	127	257	391	533	688	861	1,066
20	127	257	391	533	687	860	1,064
21	127	257	391	532	686	859	1,063
22	127	256	390	532	686	858	1,061
23	127	256	390	532	685	858	1,060
24	127	256	590	531	685	857	1,059
25	127	256	390	531	684	856	1,058
26	127	256	390	531	684	856	1,058
27	127	256	389	531	684	855	1,057
28	127	256	389	530	683	855	1,056
29	127	256	389	530	683	854	1,055
30	127	256	389	530	683	854	1,055
40	126	255	388	529	681	851	1,050
60	126	254	387	527	679	848	1,046
120	126	254	386	526	677	845	1,041
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,030
$n-1$ / β	0,1	0,2	0,3	0,4	0,5	0,6	0,7

Table 5

pour différentes valeurs de β et de $n-1$

0,8	0,9	0,95	0,98	0,99	0,999	β / $n-1$
3,08	6,31	12,71	31,8	63,7	636,6	1
1,886	2,92	4,30	6,96	9,92	31,6	2
1,638	2,35	3,18	4,54	5,84	12,94	3
1,533	2,13	2,77	3,75	4,60	8,61	4
1,476	2,02	2,57	3,36	4,03	6,86	5
1,440	1,943	2,45	3,14	3,71	5,96	6
1,415	1,895	2,36	3,00	3,50	5,40	7
1,397	1,860	2,31	2,90	3,36	5,04	8
1,383	1,833	2,26	2,82	3,25	4,78	9
1,372	1,812	2,23	2,76	3,17	4,59	10
1,363	1,796	2,20	2,72	3,11	4,49	11
1,356	1,782	2,18	2,68	3,06	4,32	12
1,350	1,771	2,16	2,65	3,01	4,22	13
1,345	1,761	2,14	2,62	2,98	4,14	14
1,341	1,753	2,13	2,60	2,95	4,07	15
1,337	1,746	2,12	2,58	2,92	4,02	16
1,333	1,740	2,11	2,57	2,90	3,96	17
1,330	1,734	2,10	2,55	2,88	3,92	18
1,328	1,729	2,09	2,54	2,86	3,88	19
1,325	1,725	2,09	2,53	2,84	3,85	20
1,323	1,721	2,08	2,52	2,83	3,82	21
1,321	1,717	2,07	2,51	2,82	3,79	22
1,319	1,714	2,07	2,50	2,81	3,77	23
1,318	1,711	2,06	2,49	2,80	3,74	24
1,316	1,708	2,06	2,48	2,79	3,72	25
1,315	1,706	2,06	2,48	2,78	3,71	26
1,314	1,703	2,05	2,47	2,77	3,69	27
1,313	1,701	2,05	2,47	2,76	3,67	28
1,311	1,699	2,04	2,46	2,76	3,66	29
1,310	1,697	2,04	2,46	2,75	3,65	30
1,303	1,684	2,02	2,42	2,70	3,55	40
1,296	1,671	2,00	2,39	2,66	3,46	60
1,289	1,658	1,980	2,36	2,62	3,37	120
1,282	1,645	1,960	2,33	2,58	3,29	∞
0,8	0,9	0,95	0,98	0,99	0,999	$n-1$ / β

Table des logarithmes de base 2 des nombres entiers de 1 à 100

x	log x	x	log x	x	log x
1	0,00000	36	5,16993	71	6,14975
2	1,00000	37	5,20945	72	6,16992
3	1,58496	38	5,24793	73	6,18982
4	2,00000	39	5,28540	74	6,20945
5	2,32193	40	5,32193	75	6,22882
6	2,58496	41	5,35755	76	6,24793
7	2,80735	42	5,39232	77	6,26679
8	3,00000	43	5,42626	78	6,28540
9	3,16993	44	5,45943	79	6,30378
10	3,32193	45	5,49185	80	6,32193
11	3,45943	46	5,52356	81	6,33985
12	3,58496	47	5,55459	82	6,35755
13	3,70044	48	5,58496	83	6,37504
14	3,80735	49	5,61471	84	6,39232
15	3,90689	50	5,64386	85	6,40939
16	4,00000	51	5,67242	86	6,72626
17	4,08746	52	5,70044	87	6,44294
18	4,16993	53	5,72792	88	6,45943
19	4,24793	54	5,75489	89	6,47573
20	4,32193	55	5,78136	90	6,49185
21	4,39232	56	5,80735	91	6,50779
22	4,45943	57	5,83289	92	6,52356
23	4,52356	58	5,85798	93	6,53916
24	4,58496	59	5,88264	94	6,55459
25	4,64386	60	5,90689	95	6,56986
26	4,70044	61	5,93074	96	6,58496
27	4,75489	62	5,95420	97	6,59991
28	4,80735	63	5,97728	98	6,61471
29	4,85798	64	6,00000	99	6,62936
30	4,90689	65	6,02237	100	6,64386
31	4,95420	66	6,04439		
32	5,00000	67	6,06609		
33	5,04439	68	6,08746		
34	5,08746	69	6,10852		
35	5,12928	70	6,12928		

Table 7

Table des valeurs de la fonction $\eta(p) = -p \log_2 p$

p	η (p)	Δ	p	η (p)	Δ
0	0	664	0,50	0,5000	—46
0,01	0,0664	464	0,51	4954	—48
0,02	1128	390	0,52	4906	—52
0,03	1518	340	0,53	4854	—54
0,04	1858	303	0,54	4800	—56
0,05	2161	274	0,55	4744	—59
0,06	2435	251	0,56	4685	—62
0,07	2686	229	0,57	4623	—65
0,08	2915	211	0,58	4558	—67
0,09	3126	196	0,59	4491	—69
0,10	3322	181	0,60	4422	—72
0,11	3503	168	0,61	4350	—74
0,12	3671	155	0,62	4276	—77
0,13	3826	145	0,63	4199	—78
0,14	3971	134	0,64	4121	—81
0,15	4105	125	0,65	4040	—83
0,16	4230	116	0,66	3957	—86
0,17	4346	107	0,67	3871	—87
0,18	4453	99	0,68	3784	—90
0,19	4552	92	0,69	3694	—92
0,20	4644	84	0,70	3602	—94
0,21	4728	78	0,71	3508	—96
0,22	4806	71	0,72	3412	—98
0,23	4877	67	0,73	3314	—99
0,24	4941	59	0,74	3215	—102
0,25	5000	53	0,75	3113	—104
0,26	5053	47	0,76	3009	—106
0,27	5100	42	0,77	2903	—107
0,28	5142	37	0,78	2796	—109
0,29	5179	32	0,79	2687	—112
0,30	5211	27	0,80	2575	—113
0,31	5238	22	0,81	2462	—114
0,32	5260	18	0,82	2348	—117
0,33	5278	14	0,83	2231	—119
0,34	5292	9	0,84	2112	—120
0,35	5301	5	0,85	1992	—121
0,36	5306	1	0,86	1871	—123
0,37	5307	—2	0,87	1748	—125
0,38	5305	—7	0,88	1623	—127
0,39	5298	—10	0,89	1496	—128
0,40	5288	—14	0,90	1368	—130
0,41	5274	—18	0,91	1238	—131
0,42	5256	—20	0,92	0,1107	—133
0,43	5236	—24	0,93	0,0974	—135
0,44	5210	—26	0,94	839	—136
0,45	5184	—29	0,95	703	—138
0,46	5153	—33	0,96	565	—139
0,47	5120	—37	0,97	426	—140
0,48	5083	—40	0,98	286	—142
0,49	5043	—43	0,99	144	—144
0,50	0,5000		1.00	0	

Table 8

Valeurs de $P_m = \frac{a^m}{m!} e^{-a}$ (répartition de Poisson)

m	a=0,1	a=0,2	a=0,3	a=0,4	a=0,5	a=0,6	a=0,7	a=0,8	a=0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0019	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0001	0,0002	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003

m	a=1	a=2	a=3	a=4	a=5	a=6	a=7	a=8	a=9	a=10
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,0613	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0.0378
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0001	0,0037	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901
8		0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0.1126
9		0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,1318	0,1251
10			0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11			0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12			0,0001	0,0006	0,0034	0,0126	0,0263	0,0481	0,0728	0,0948
13				0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14				0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15					0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16						0,0003	0,0014	0,0045	0,0109	0,0217
17						0,0001	0,0006	0,0021	0,0058	0,0128
18							0,0002	0,0009	0,0029	0,0071
19							0,0001	0,0004	0,0014	0,0037
20								0,0002	0,0006	0,0019
21								0,0001	0,0003	0,0009
22									0,0001	0,0004
23										0,0002
24										0,0001

INDEX ALPHABETIQUE DES MATIÈRES

TABLE DES MATIÈRES